



Interval-valued fuzzy regression: Philosophical and methodological issues

Reda Boukezzoula, Didier Coquin

► To cite this version:

Reda Boukezzoula, Didier Coquin. Interval-valued fuzzy regression: Philosophical and methodological issues. Applied Soft Computing, 2021, pp.107145. 10.1016/j.asoc.2021.107145 . hal-03138248

HAL Id: hal-03138248

<https://hal.science/hal-03138248>

Submitted on 13 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Interval-Valued Fuzzy Regression: Philosophical and Methodological Issues

*Reda Boukezzoula, Didier Coquin

Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance - LISTIC
Polytech Annecy-Chambéry, 5 chemin de bellevue, Annecy-le-vieux, 74940 Annecy, Université Savoie Mont Blanc, France.

Abstract

This paper revisits interval-valued fuzzy regression and proposes a new unified framework to address interval-valued type-1 and type-2 fuzzy regression models. The paper focuses on two main objectives. First, some philosophical and methodological reflections about interval-valued type-1 fuzzy regression (IV-T1FR) and interval-valued type-2 fuzzy regression (IV-T2FR) are discussed and analyzed. These reflections aim at positioning fuzzy regression to avoid misinterpretations that may sometimes lead to erroneous or ambiguous considerations in practical applications. Consequently, the interest, relevance, representativeness and typology of interval-valued fuzzy regression are established. Therefore, IV-T1FR generalizes conventional interval regression (CIR) and increases its specificity. However, if the IV-T1FR can fit fuzzy data, then its formalism is not able to address the uncertainty phenomenon in the IV-T1FS representation. In this context, the IV-T2FR can be regarded as an uncertain IV-T1FR, i.e., a generalization of the IV-T1FR in an uncertain environment. Second, a new unified methodology to address fuzzy regression models using the concepts of gradual intervals (GIs) and thick gradual intervals (TGIs) is proposed. The proposed view allows handling regression models *via* an extension of the standard interval arithmetic (SIA) —initially proposed for conventional intervals (CIs)— to GIs and TGIs. The originality of the proposed approach resides in the fact that all the CIR methodologies can be extended to the IV-T1FR methods. Furthermore, all the IV-T1FR methodologies can be extended to the IV-T2FR framework. Our view does not depend on the model shape and preserves the flexibility and rigor of SIA computations in the propagation of fuzzy quantities through regression models. The proposed concepts are validated using illustrative examples.

Key words: Conventional Interval Regression (CIR), Gradual intervals and Thick Gradual Intervals, Interval-Valued Type-1 Fuzzy Regression (IV-T1FR), Interval-Valued Type-2 Fuzzy Regression (IV-T2FR), Uncertainty.

1. Introduction

In data representation and analysis, the usual statistical regression (SR) approach aims to determine an input-output (IO) relationship between variables (dependent and independent) where deviations between the observed and estimated values are assumed to be due to random errors. In the literature, SR approaches—based on probabilistic concepts—have been largely exploited where the observations are treated as random variables using probability distributions [20]. In these methods, the verification of their statistical assumptions requires many observations. However, the observations are sometimes insufficient in quantity and/or in quality. Furthermore, some pieces of information may be missing due to the additive costs that would result from their acquisition. Moreover, in some applications, data imperfection is not induced by randomness but by the presence of empirical information, conflict data and/or vagueness assessment [6][20][24]. Usually, probabilistic techniques can impose complex computations, especially in nonlinear problems. In such a context where the probability density estimation is difficult, one should move toward more flexible methods that can relax the strict requirements on the statistical properties of the manipulated data. Therefore, the interval-based approach can represent a possible alternative.

* Corresponding author. E-mail address: reda.boukezzoula@univ-smb.fr.
Université Savoie Mont Blanc, France, Tel: (+33) 450096526. Fax: (+33) 450096559

Unlike SR methods, when CIR approaches are considered [7][12][41][44], the uncertainty associated with the manipulated data is only represented by CIs (sets of likelihood). This view allows easier computation with fewer statistical hypotheses. Usually, and rightly, the CI representation can be considered relatively poor and less specific than the probabilistic representation. Therefore, in SR approaches, a variable is represented by a probability density that integrates two kinds of information: the support, which surely contains the variable, and the distribution of this variable on this support. From practical perspectives, when amply rich information is available to suitably estimate probability densities of the variables and their independence/dependence, the probabilistic view is better and richer than the interval-based approach. However, CI methods, which can address insufficient data, allow more flexible manipulations in complex and nonlinear computations by avoiding strict statistical requirements that are not always verified in practice. From methodological perspectives, the statistical and interval-based approaches are not in opposition but rather complementary. Furthermore, there are many examples of applications with intrinsically interval-valued data (the variables do not take crisp values but CI values) [35][38]. In this framework, standard statistical methods for single-valued data are not adapted to properly take into account these interval data.

Based on the CI representation, many CIR approaches were proposed (e.g., [7][10][23][28][41][44]) where the SIA [28][66][57] is used to formalize the IO mapping as an interval-valued function. It is important to emphasize that the support of a distribution resulting from a probabilistic computation is always included in the CI obtained by SIA on these supports [57]. Conceptually, the CIR process is a fitting problem in which inferential concepts are not considered and no probabilistic distribution assumptions are established on the manipulated data and models. This view is justified by the fact that the finality and the use of SR and CIR approaches are distinct. The CIR provides guaranteed bounds, while the SR provides confidence intervals. Moreover, the CIR privileges the flexibility and guaranteed results of SIA for elaborating regression models with realistic and softer hypotheses.

Usually, SR approaches address data according to an epistemic view (the uncertainty of point-value quantities). Unlike SR methods, CIR approaches can address epistemic and ontic interpretations [22][25][32][61]. In the epistemic approach, a CI is regarded as a set of possible values, one of which is the true one (the CI exhibits partial or incomplete information) [25]. In the ontic view, a CI can be regarded as the precise description of a set-valued entity [25]. This dichotomy, which will be discussed later, is essential in understanding the essence of the CIR regression, its extensions, and its practical application.

In the CI representation, any value in a CI $[a]$ is possible, and any value outside $[a]$ is impossible. To remedy this lack of specificity and permit the representation of the progressive belonging to a CI, fuzzy regression has been envisioned. The fuzzy regression philosophy is based on the concept of fuzzy sets (FSs) for fitting the IO data. A usual FS is known as a type-1 FS (T1FS). The term type-1 is used to distinguish it from a type-2 FS (T2FS) [63][65][82]. Due to their simplicity of computations, interval-valued (type-1 and type-2) FSs have been largely exploited. For example, an interval-valued T1FS (IV-T1FS) is a convex T1FS where all its α -cuts are CIs [31] (an IV-T1FS is a stack of nested CIs defined by the α -cuts). Conceptually, interval-valued type-1 fuzzy regression (IV-T1FR) has often been regarded as a CIR *via* the concept of α -cuts [10][18][20]. Therefore, the IV-T1FR equips the CIR by a vertical dimension to exhibit the distribution of the variables. These distributions are not probability distributions but possibility distributions (an IV-T1FS can be regarded as a possibility distribution). The IV-T1FR—which generalizes the CIR—is a non-SR approach and is not based on probability theory but rather on possibility theory.

The distinction between ontic vs. epistemic CIs remains valid for IV-T1FSs and IV-T2FSs, and it is useful for understanding the essence of fuzzy regression and its methodology. The usefulness of this distinction is not limited to fuzzy regression and can be interesting in other fields, such as decision-making problems. Therefore, in complex and uncertain environments, decision makers are often unable to provide crisp values for their evaluations and/or judgments. They often use some degree of imprecision and/or uncertainty to formulate their human subjective judgments. To address this situation, IV-T1FSs and IV-T2FSs have been

substantially exploited in decision-making problems. In this decision-making framework, comparability and ordering relations between the interval-valued fuzzy quantities must be done according to the meaning attributed to them (ontic or epistemic) to avoid incoherent operations. This dichotomy between ontic and epistemic is necessary to build coherent preference relations. Furthermore, the final ranking of alternatives based on incoherent preference relations may lead to incorrect and ambiguous decisions (see [68] for a comparison of uncertainty data using epistemic and ontic views in decision-making problems).

The IV-T1FR analysis aims at finding a regression model that fits the observed IV-T1FS data within a specified criterion. In the literature, two different approaches have been proposed depending on the fitting criterion used. The first is the possibilistic approach (e.g., [12][43][50][73][74][75]). The second is the least-square (LS) approach (e.g., [5][30][33][51][52]), which extends the LS criterion to the fuzzy case. To be able to extend the SIA arsenal to IV-T1FSs, an IV-T1FS is regarded as a GI, i.e., a pair of lower and upper gradual numbers (bounds), which are named left and right profiles [11][12][39]. Therefore, *via* GIs, a revision and a new interpretation of the IV-T1FR is proposed.

In the literature, IV-T1FR has been successfully applied to countless applications in different fields (see [24] for a good survey). However, the IV-T1FR is not equipped to contend with IV-T2FSs [36][37][64] and/or situations where the shape of IV-T1FSs is not precisely known or is hidden for confidentiality purposes. In this framework, IV-T2FR can be considered an alternative. Therefore, IV-T2FR can be regarded as an uncertain version of IV-T1FR. Although IV-T1FR has been extensively studied and analyzed, only a few works have studied IV-T2FR (see [2][53][69][79]). While these methods can address type-2 regression, they are premature and partially represent IV-T2FSs. Furthermore, these methods involve many parameters, and their associated computational mechanism is expensive and difficult to generalize to any regression model. This work aims to prove that IV-T2FR can be naturally extended from CIR and IV-T1FR according to the concept of TGIs [14]. A TGI is a new representation of an IV-T2FS, which is based on the notion of thick intervals (TIs) [14][29]. A TI is a new kind of CI in which its bounds are also CIs.

The motivation of this paper is twofold. First, the philosophical and methodological principles of IV-T1FR and IV-T2FR are discussed and analyzed. We show that the IV-T1FR is a generalization of the CIR in which the interval specificity has been enriched through possibility distributions. Based on the TGI concept, it has been shown that IV-T2FR is an uncertain IV-T1FR. Second, this paper proposes a unified formalism to address IV-T1FR and IV-T2FR. This work demonstrates and states that the propagation of the information in interval-valued fuzzy regression models is achieved by an extension of the SIA—initially proposed for CIs—to GIs for the IV-T1FR and to TGIs for the IV-T2FR. The originality of the proposed approach resides in the fact that all the CIR methodologies can be extended to the IV-T1FR methods where the IO variables and/or parameters are represented by GIs. Furthermore, the IV-T1FR can be extended to the IV-T2FR framework where the inputs, outputs and/or the parameters are represented by TGIs. Regarding the IV-T1FR and IV-T2FR methods that are published in the literature, our vision presents a unified methodology that does not depend on the model shape and preserves the flexibility and rigor of SIA computations in the fuzzy regression framework.

The remainder of this paper is organized as follows: Section 2 is devoted to some preliminaries and notations. Section 3 discusses the representativeness of CI models and interval-valued fuzzy models. The typology of the interval-valued fuzzy regression is given in section 4. Section 5 is devoted to the methodology of implementation of the interval-valued fuzzy regression. Section 6 is dedicated to computational examples according to IV-T1FR and IV-T2FR. Remarks, limitations and discussion of our approach are given in section 7. Finally, conclusions are presented in section 8.

2. Preliminaries and notations

For the sake of rigor and clarity, let us define the basic concepts used in this paper. Some of these concepts have been published in [8][11][12][14]. For more details, the reader is invited to consult these references. Table 1 shows the list of abbreviations used in the paper.

Abbreviations	Full meaning	Abbreviations	Full meaning
SR	statistical regression	T2FS	type-2 fuzzy set
IO	input-output	IV-T1FS	interval-valued type-1 fuzzy set
CI	conventional interval	IV-T2FS	interval-valued type-2 fuzzy set
SIA	standard interval arithmetic	CIR	conventional interval regression
GI	gradual interval	IV-T1FR	interval-valued type-1 fuzzy regression
TI	thick interval	IV-T2FR	interval-valued type-2 fuzzy regression
TGI	thick gradual interval	PM	parametric model
FS	fuzzy set	NPM	nonparametric models
T1FS	type-1 fuzzy set	AIC	Akaike information criterion

Table 1: List of abbreviations

2.1. Conventional intervals (CIs) and gradual intervals (GIs)

A real interval $[a]$ is defined as a closed compact and bounded subset of \Re such that:

$$[a] = [a^-, a^+] = \{a \in \Re \mid a^- \leq a \leq a^+\} \quad (1)$$

In (1), the real numbers a^- and a^+ are the endpoints (the lower and upper bounds, respectively) of $[a]$. The set $\mathbb{I} = \{[a^-, a^+] \mid a^- \leq a^+; a^-, a^+ \in \Re\}$ denotes the set of CIs. Representing information a by a CI $[a]$ means that any value in $[a]$ is possible and any value outside is impossible. To introduce the concept of progressive belonging to $[a]$, the concept of GIs has been proposed [11][12][39]. The GI enriches the CI representation and improves its specificity. Therefore, the CI $[a]$ is then transformed into a GI $[a(\lambda)]$ if its bounds are GNs [39]. A GN is a number that is parameterized by a degree of pertinence or flexibility $\lambda \in [0, 1]$. As for a CI, a GI is represented by its bounds, which are called left and right profiles, i.e., $[a(\lambda)] = [a^-(\lambda), a^+(\lambda)]$, where $a^-(\lambda) \leq a^+(\lambda)$. The GI $[a(\lambda)]$ is an IV-T1FS if its profiles $a^-(\lambda)$ and $a^+(\lambda)$ are injective and nondecreasing and nonincreasing, respectively [11][12]. Therefore, an IV-T1FS is a special case of a GI. In a gradual framework, the monotonicity constraint associated with the bounds $a^-(\lambda)$ and $a^+(\lambda)$ is not imposed. Therefore, if $a^-(\lambda)$ and $a^+(\lambda)$ are, respectively, nondecreasing and nonincreasing, $[a(\lambda)]$ is said to be monotone (consonant) and is regarded as an IV-T1FS. In contrast, $[a(\lambda)]$ is said to be pure GI and cannot be characterized by an IV-T1FS. For example, the GI $[a(\lambda)]$ in Fig. 1.a is consonant (monotone) and can be regarded as an IV-T1FS. However, $[b(\lambda)]$ in Fig. 1.b is a pure GI. The SIA operations on CIs can be extended to GIs where all the CIs are represented by GIs. These operations are defined in Appendix A (see also [11][14]). Similar to a CI $[a]$ where its radius is given by $R([a]) = (a^+ - a^-)/2$, the radius of a GI $[a(\lambda)]$ can be expressed by $R([a(\lambda)]) = (a^+(\lambda) - a^-(\lambda))/2$.

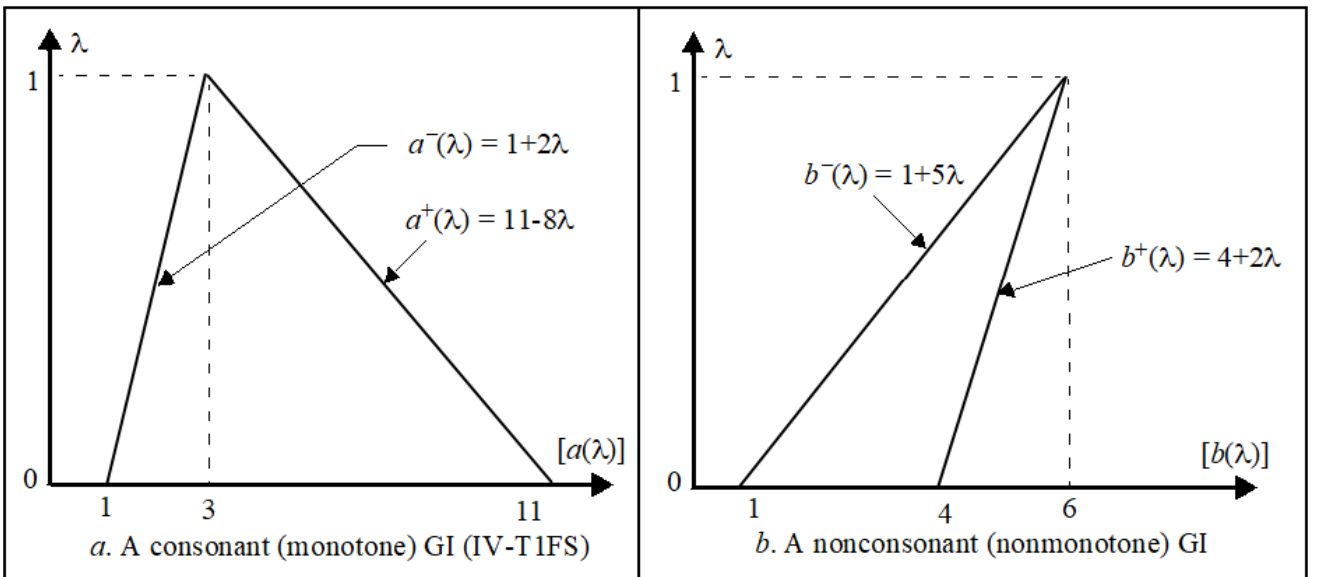


Fig. 1: Consonant (IV-T1FS) and nonconsonant (pure) GIs

2.2. Thick intervals (TIs) and thick gradual intervals (TGIs)

A TI, denoted $\llbracket a \rrbracket$, is used for representing CIs whose bounds are uncertain and are also represented by CIs [14][29]. By analogy with a CI, a TI is a subset of \mathbb{I} and can be expressed by (refer to Fig. 2):

$$\llbracket a \rrbracket = \llbracket [a^-], [a^+] \rrbracket = \{[a] = [a^-, a^+] \in \mathbb{I} \mid a^- \in [a^-] \text{ and } a^+ \in [a^+]\} \quad (2)$$

In (2), the bounds $[a^-]$ and $[a^+]$ are two CIs that contain the uncertain lower bound $a^- \in \mathfrak{R}$ and the uncertain upper bound $a^+ \in \mathfrak{R}$, respectively. The TI $\llbracket a \rrbracket$ exhibits all the CIs $[a^-, a^+]$, where $a^- \in [a^-]$ and $a^+ \in [a^+]$ (see Fig. 2). A CI is a TI for which the bounds are exactly known, i.e., $[a] = [a^-, a^+] = \llbracket [a^-, a^-], [a^+, a^+] \rrbracket$.

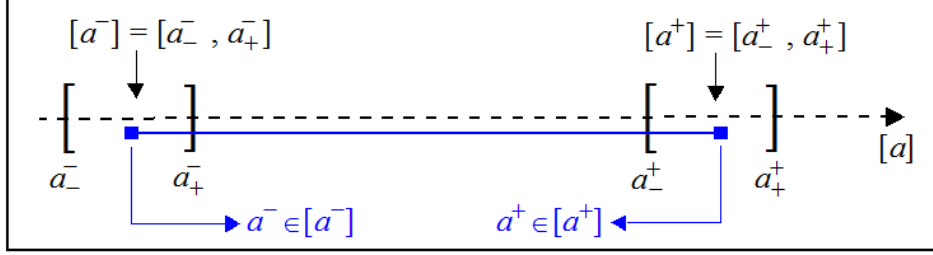


Fig. 2: Representation of a thick interval (TI) $\llbracket a \rrbracket$

The SIA operations over CIs can be directly extended to TIs [14][29] (see Appendix A). The radius of a TI is defined by the radii of its CI bounds, i.e., $R(\llbracket a \rrbracket) = R([a^-]) + R([a^+])$. Furthermore, the specificity of TIs has been enriched by their extension to TGIs. The concept of TGIs, which has been proposed in [14], is useful for the manipulation of IV-T2FSs. Therefore, a GI $[a(\lambda)]$ becomes a TGI $\llbracket a(\lambda) \rrbracket = \llbracket [a^-(\lambda)], [a^+(\lambda)] \rrbracket$, where the bounds $[a^-(\lambda)]$ and $[a^+(\lambda)]$ are GIs. A TGI $\llbracket a(\lambda) \rrbracket$ is defined by:

$$\llbracket a(\lambda) \rrbracket = \llbracket [a^-(\lambda)], [a^+(\lambda)] \rrbracket = \{[a(\lambda)] = [a^-(\lambda), a^+(\lambda)] \in \mathbf{GI} \mid a^-(\lambda) \in [a^-(\lambda)] \text{ and } a^+(\lambda) \in [a^+(\lambda)]\} \quad (3)$$

In (3), \mathbf{GI} represents the set of GIs. The TGI $\llbracket a(\lambda) \rrbracket$ represents an uncertain GI $[a(\lambda)]$ (IV-T1FS) where its profiles $a^-(\lambda)$ and $a^+(\lambda)$ are uncertain and located in their likelihood GIs $[a^-(\lambda)]$ and $[a^+(\lambda)]$, respectively (see Fig. 3). By analogy with the GI formalism, $[a^-(\lambda)]$ and $[a^+(\lambda)]$ are, respectively called left and right GIs. According to Fig. 3, the bounds $[a^-(\lambda)]$ and $[a^+(\lambda)]$ are pure GIs and cannot be represented by IV-T1FSs. As detailed in [14], through the TGI concept, a new interpretation of IV-T2FSs is possible. Therefore, an IV-T2FS can be regarded as a TGI $\llbracket a(\lambda) \rrbracket = \llbracket [a^-(\lambda)], [a^+(\lambda)] \rrbracket$ (and *vice versa*), which is composed of two pure GIs $[a^-(\lambda)]$ and $[a^+(\lambda)]$: one for representing the left part and the other the right part, respectively (see Fig. 3). For more details on the equivalence between IV-T2FSs and TGIs, see [14]. Moreover, the SIA operations on TIs have been extended to TGIs where all the TIs are replaced by TGIs. These SIA operations are defined in Appendix B.

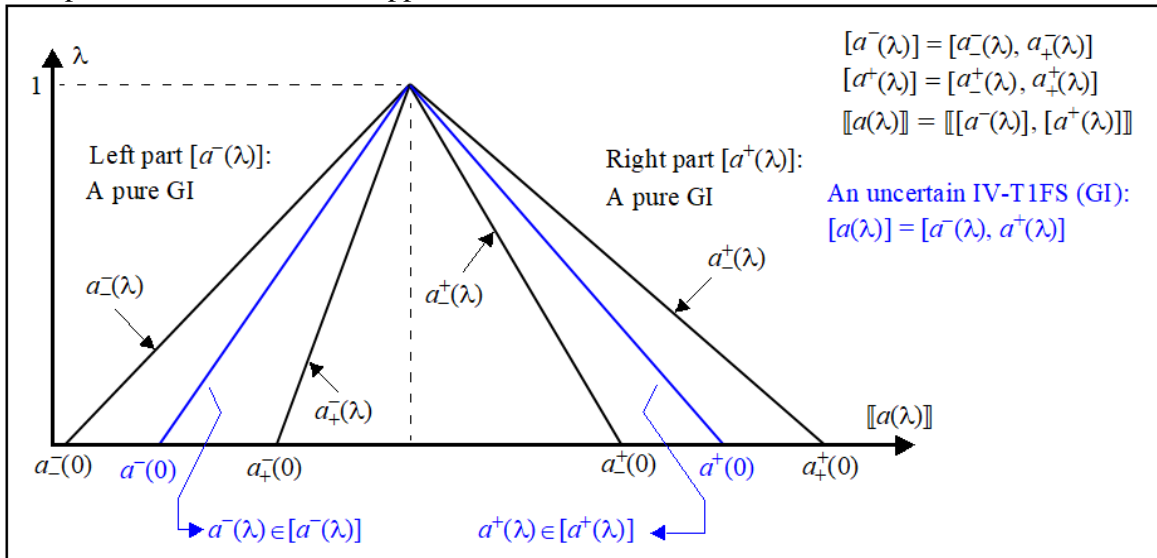


Fig. 3: Representation of a TGI $\llbracket a(\lambda) \rrbracket$

3. Interval-valued fuzzy model representativeness

3.1. CI models: ontic and epistemic interpretations?

In CIR models, the inputs, outputs and/or parameters are represented by CIs [12][23][44]. Through these models, two different semantics of the manipulated CIs can be distinguished: the imprecise description of a point-valued quantity (epistemic view) and the precise description of a set-valued entity (ontic view) [25]. This differentiation is indispensable in understanding the philosophy of CIR and its implementation.

In the epistemic view, a CI is regarded as set-valued data points, which represents an imprecise observation of the quantity of interest. This uncertainty represents the idea of partial or incomplete information. Therefore, it may be described using a set of possible values, one of which is the true value. Simply, an epistemic CI is regarded as a set of disjunctive elements [25][61]. This representation is always represented by a disjunction of mutually exclusive elements, only one of which is the true value of the considered quantity. Furthermore, in the majority of experimental scenarios, CIs are epistemic and represent exact but imprecise measurements. The experimental data are not precisely measured or are hidden for confidentiality purposes. Therefore, only CIs that contain the true values are available (the measurements are precise, but they are instantiated in their likelihood CIs). For example, if the measurement of a physical quantity (temperature θ , for example) led to stating that θ is between 18°C and 24°C, then we can write $\theta \in [18, 24]$. The CI $[\theta] = [18, 24]$ represents a set of disjunctive mutually exclusive elements, one of which is the true value of θ (refer to Fig. 4). In a discrete and integer context, $\theta \in [18, 24]$ is understood as:

$$\theta \text{ is between 18 and 24} \Leftrightarrow \theta \in [\theta] = [18, 24] \Leftrightarrow \theta = 18 \vee 19 \vee \dots \vee 24$$

Unlike the epistemic view, a CI may be conjunctive and can be called an ontic CI. In this situation, the CI is used for representing some precise and compact entity (the information is intrinsically given in the form of CIs) [25][61]. From a philosophical point of view, an ontic interval is regarded as a set-valued data, which is interpreted as the observation of an entity that can be regarded as “imprecise or uncertain” only in the sense that the observed data are not single-valued points but set-valued data. Therefore, an ontic CI is set-valued by nature and there are no distinguished elements in the observed set. Furthermore, time intervals are typical applications of ontic intervals (lifetime of a person, lifespan of an electronic component, etc.). For example, let us take the example of the life of a person. Say, Joseph Fourier lived from 1768 to 1830. At every time point in the CI [1768, 1830], Fourier was alive, but there is no special distinguished time point within this period (see Fig. 4). In this context, if we denote the lifetime of Fourier by v , in a discrete and integer context, it is possible to note:

$$v = [v] = [1768, 1830] \Leftrightarrow v = 1768 \wedge 1769 \wedge \dots \wedge 1830$$

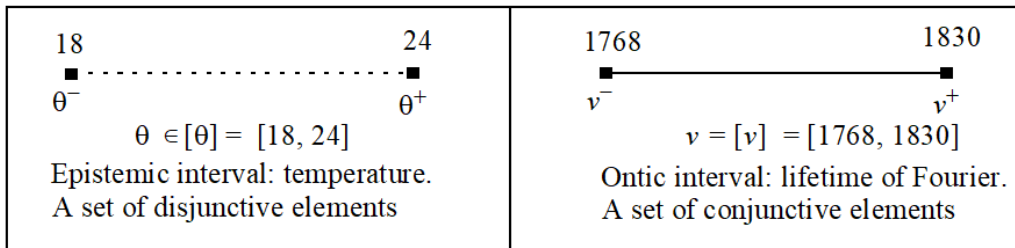


Fig 4: Epistemic and ontic interpretations

The distinction between conjunctive (ontic) and disjunctive (epistemic) CIs can be extended to address IV-T1FSs and IV-T2FSs. This point will be discussed later in the paper. When CI data are manipulated through regression models, it is important to know and specify the semantics associated with the used data (ontic or epistemic). This categorization can improve the relevance of the used method and avoid misinterpretations that may sometimes lead to ambiguous considerations. Therefore, in practical applications, it can sometimes be surprising to apply methods that are designed in an ontic fashion to situations of epistemic data and *vice versa*. Furthermore, as mentioned in [7], in the majority of the experimental scenarios, the intervals (CIs, IV-T1FSs, IV-T2FSs) are of an epistemic nature. In this context, epistemic regression approaches can be considered the most relevant.

3.2. CI-based models *versus* probabilistic-based models

Most of the regression methodologies are statistical, and they are based on probabilistic concepts. These techniques take advantage of the fact that probability distributions govern the manipulated data. They involve treating observations as random variables in probability distributions and computing something about these distributions. Let us give a simple example to illustrate this approach. Consider the three random variables in_1 , in_2 and out of \Re . These variables are linked through the simple linear model $out = in_1 + in_2$, where out is the output (dependent variable) and in_1 and in_2 are the inputs (independent variables). In the probabilistic view, some statistical properties on the inputs in_1 and in_2 are needed for deducing those of the output out . Let us assume that the distributions (π) of in_1 and in_2 have been identified as uniform densities, such as $\pi(in_1) = 1/5$ over $[0, 5]$ and $\pi(in_2) = 1/2$ over $[1, 3]$. Since the inputs are independent variables, the density of the couple (in_1, in_2) is uniform, i.e., $\pi(in_1, in_2) = (1/5) \times (1/2) = 1/10$ over the box $[0, 5] \times [1, 3]$. Hence, the probability density function of out is given by the following expression (see Fig. 5):

$$\pi(out) = \int_{\Re} \pi_{in_1}(t) \cdot \pi_{in_2}(out - t) dt = \begin{cases} 0, & \text{if } out \leq 1 \\ (out - 1)/10, & \text{if } out \in [1, 3] \\ 1/5, & \text{if } out \in [3, 6] \\ (8 - out)/10, & \text{if } out \in [6, 8] \\ 0, & \text{if } out \geq 8 \end{cases} \quad (4)$$

The computation of $\pi(out)$ can be more complex when the linear model becomes nonlinear. Furthermore, the independence property between the variables is often assumed. However, in some applications, this information is not available, and in most cases, it does not correspond to reality, particularly when physical sensors are used. In such situations, the CI representation can represent a possible alternative.

Unlike probabilistic methods, when considering an interval-based approach, the uncertainty associated with in_1 and in_2 are only given by their CIs of likelihood. For example, where the inputs are acquired with uncertainty and they are associated with their CIs $in_1 \in [in_1] = [0, 5]$ and $in_2 \in [in_2] = [1, 3]$, the output out is:

$$out \in [out] = [in_1] + [in_2] = [0, 5] + [1, 3] = [1, 8] \quad (5)$$

In this context, the statistical assumptions are relaxed, and the uncertainty is traduced by a CI. Therefore, representing the variable in_1 by a CI $[in_1]$ (called the likelihood set) means that $[in_1]$ contains in_1 and nothing else, i.e., this representation does not assume the probability distribution of in_1 to be uniform. For instance, the output out is described by the CI $[1, 8]$ but is not uniform on this CI.

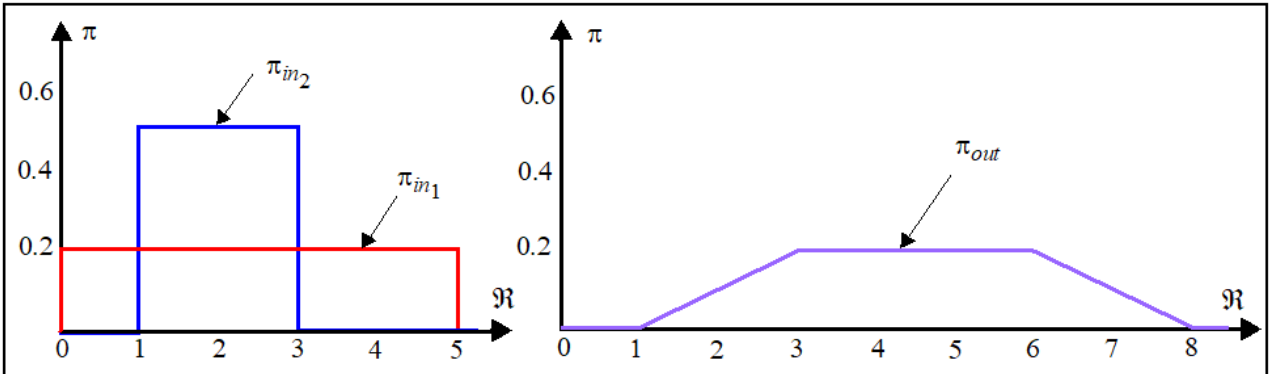


Fig 5: Probability densities of variables in_1 , in_2 and out

More generally, the CI approach allows more flexible computations where strict requirements on the statistical properties of the manipulated variables are relaxed. However, and as discussed in the introduction, the CI view presents a lack of specificity compared to a probabilistic approach. An interesting idea would be to increase the specificity of the CI models. This will be discussed in the next section.

3.3. Representativeness of IV-T1FS-based models

Usually, IV-T1FS models are based on the representation of inputs, outputs and/or parameters by IV-T1Fs. This fuzzy vision aims to increase the specificity of CI models and to enrich their representation. CI-based

models can be seen as a particular case of IV-T1FS models. However, in contrast to CI representation where only a unique horizontal dimension is considered, IV-T1FSs (monotone GIs) are represented by two dimensions (horizontal and vertical). If the horizontal dimension is similar to that which is used in CIs, the vertical dimension is related to the relevance degrees and is limited to the unit interval $[0, 1]$. These degrees of relevance can be interpreted as membership degrees for fuzzy sets, degrees of possibility, degrees of truth, degrees of flexibility, etc. From methodological perspectives, IV-T1FS regression models are CI regression models where a vertical dimension is integrated. This finding is consistent with the works published in the literature where IV-T1FR has been treated as a CIR *via* the concept of α -cuts.

As for CIs, two semantic interpretations of IV-T1FS models can coexist. The first view assumes that the IV-T1FSs are intrinsically ontic (set-valued set parametrized by a degree λ). According to the concepts proposed in [32], ontic T1FSs (and IV-T1FIs) represent objects originally constructed as sets (CIs), for which the fuzzy representation is more expressive due to the gradual boundaries. The degree λ gives an evaluation of the extent to which components participate in the global entity. The second view refers to an epistemic interpretation of the fuzzy data [32] where an IV-T1FS (T1FS in general) is regarded as a possibility distribution representing incomplete information. Therefore, this possibility distribution is associated with a realization that has been only partially or imperfectly observed. The epistemic view is by far the most exploited in the literature, and this is the approach that will be used in this paper. The IV-T1FR aims to enrich the CIR to remedy its insufficiency by introducing a new type of distribution (possibility distributions). Therefore, IV-T1FR is a non-SR approach and is not based on probability distributions but on possibility distributions (IV-T1FSs can be regarded as possibility distributions).

Let us reconsider the example of subsection III.2. Now, we consider that the inputs in_1 and in_2 are still uncertain and are instantiated in their likelihood CIs to which degrees of confidence are associated. In this case, it can be written that $in_1 \in [in_1(\lambda)]$ and $in_2 \in [in_2(\lambda)]$, where $[in_1(\lambda)]$ and $[in_2(\lambda)]$ are distributions of possibility given by the IV-T1FSs (monotone GIs): $[in_1(\lambda)] = [2\lambda, 5-3\lambda]$ and $[in_2(\lambda)] = [1+\lambda, 3-\lambda]$ (see Fig. 6). As the inputs are IV-T1FSs (GIs), the model output is also an IV-T1FS (GI). This output is obtained through an addition operation between IV-T1FSs (GIs), i.e.,

$$[out(\lambda)] = [out^-(\lambda), out^+(\lambda)] = [in_1(\lambda)] + [in_2(\lambda)] = [2\lambda, 5-3\lambda] + [1+\lambda, 3-\lambda] = [1+3\lambda, 8-4\lambda] \quad (6)$$

In the IV-T1FS models, the information is propagated, computed and evaluated through the GI arithmetic [11][12]. See Appendix B for the SIA operations on GIs.

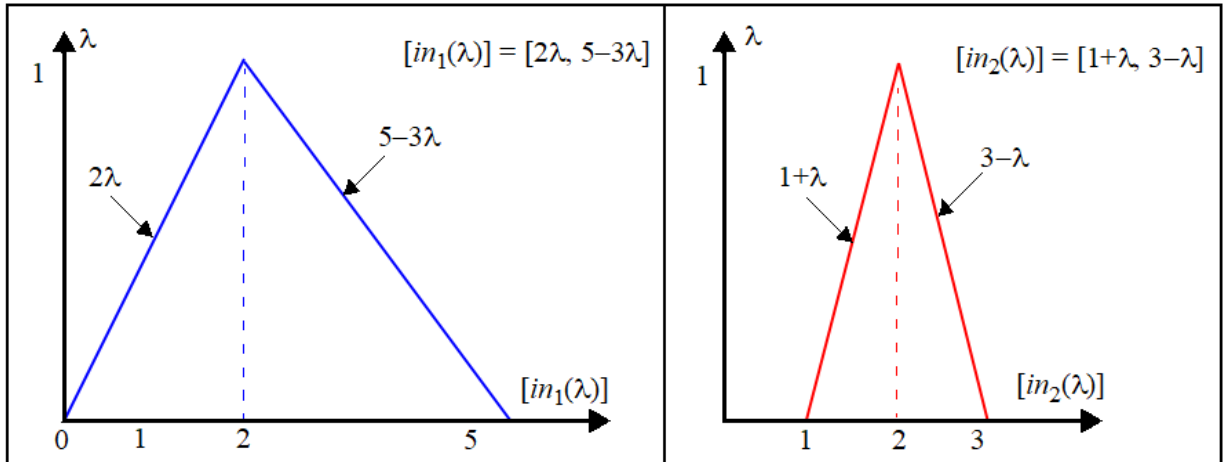


Fig 6: IV-T1FSs (possibility distributions) of $[in_1(\lambda)]$ and $[in_2(\lambda)]$

The IV-T1FS $[out(\lambda)]$ of Fig. 7 can be regarded as a set of nested confidence CIs. The degree of necessity for a CI $[out(\lambda_i)]$ to contain the value of out is $N([out(\lambda_i)]) = 1-\lambda_i$. Therefore, the confidence interval $[out(\lambda_i)]$ is interpreted as follows: « I'm certain to a degree $(1-\lambda_i)$ that out is in $[out(\lambda_i)]$ ». It is easy to show that the degree of possibility $\Pi([out(\lambda_i)]) = 1$. More generally, a set of nested confidence CIs $[out(\lambda_i)]$ with degrees of certainty $(1-\lambda_i)$ is equivalent to a distribution possibility (an IV-T1FS).

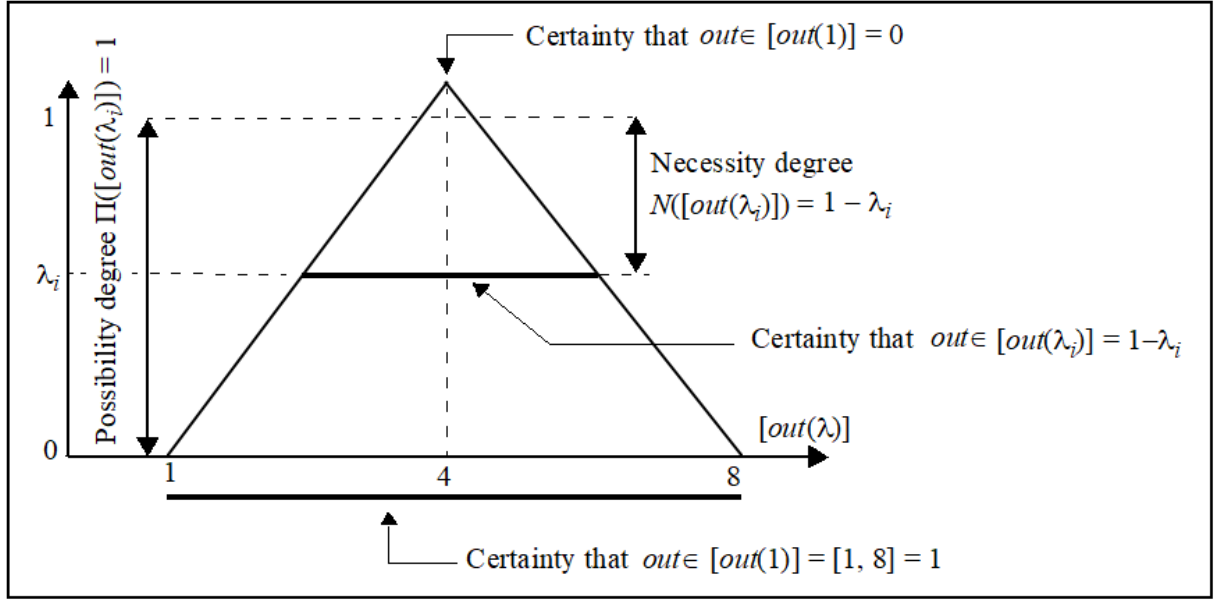


Fig 7: IV-T1FS (the possibility distribution) of $[out(\lambda)]$ and its interpretation

3.4. Representativeness of IV-T2FS-based models

To address uncertainties in T1FSs, the concept of T2FSs was introduced by Zadeh [82] as an extension of T1FSs. A T2FS is a fuzzy set whose membership grades themselves are T1FSs. Due to the computational complexity of T2FSs, some researchers have concentrated on IV-T2FSs [36][37][64]. An IV-T2FS is useful in situations where it is difficult to determine the exact shape of an IV-T1FS. Therefore, IV-T2FR aims to address IV-T1FSs (monotone GIs) whose bounds (profiles) are uncertain.

Naturally, an IV-T2FS is defined by the two IV-T1FSs—lower IV-T1FS and upper IV-T1FS—with the inclusion constraint that the lower IV-T1FS \subseteq the upper IV-T1FS. In this paper, another representation is used where an IV-T2FS is regarded and represented as a TGI [14]. This concept extends IV-T1FR to IV-T2FR, where information propagation through linear and nonlinear models is realized using SIA over TGIs. Therefore, all the inputs, outputs and/or parameters can be represented by TGIs.

The ontic and epistemic dichotomy discussed for CIs and IV-T1FSs can be extended to an IV-T2FS interpretation and meaning (refer to [16]). Therefore, an IV-T2FS can be regarded as a set of conjunctive IV-T1FSs in an ontic view. In the epistemic interpretation, an IV-T2FS is considered a set of disjunctive IV-T1FSs where the exact IV-T1FS is unknown, and only its bounds are available.

Let us reconsider the example of subsection III.3 and assume that the IV-T1FSs $[in_1(\lambda)]$ and $[in_2(\lambda)]$ of Fig. 6 are uncertain and cannot be known with certainty (the confidence intervals cannot be determined with precision or have been specified by different experts). In this case, the locations of the left and right profiles of IV-T1FSs are assumed to belong to their likelihood GIs (see Fig. 8):

$$\begin{cases} in_1^-(\lambda) \in [in_1^-(\lambda)] = [-1+3\lambda, 1+\lambda] \\ in_1^+(\lambda) \in [in_1^+(\lambda)] = [4-2\lambda, 6-4\lambda] \end{cases}, \begin{cases} in_2^-(\lambda) \in [in_2^-(\lambda)] = [0.5+1.5\lambda, 1.5+0.5\lambda] \\ in_2^+(\lambda) \in [in_2^+(\lambda)] = [2.5-0.5\lambda, 3.5-1.5\lambda] \end{cases}$$

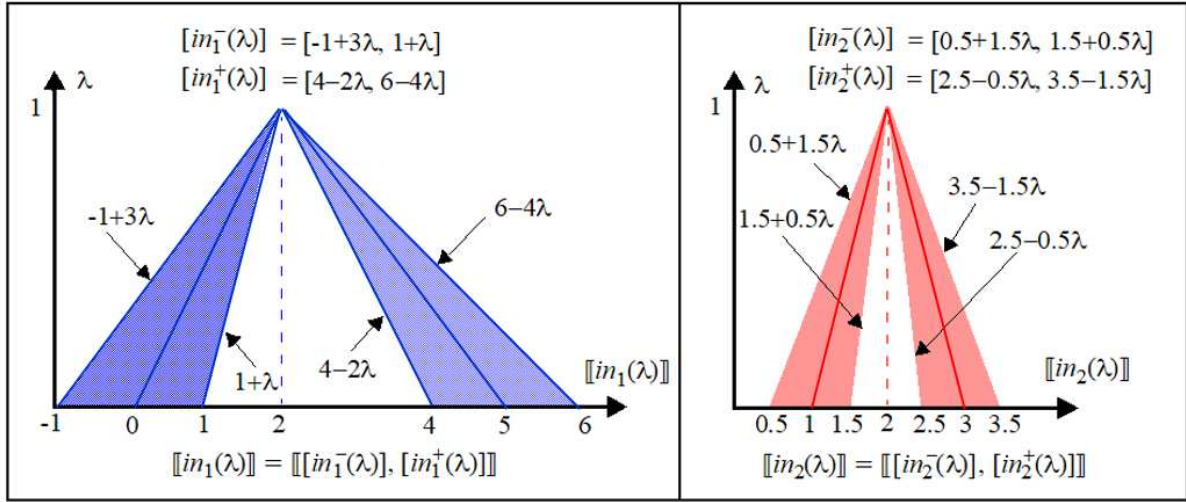
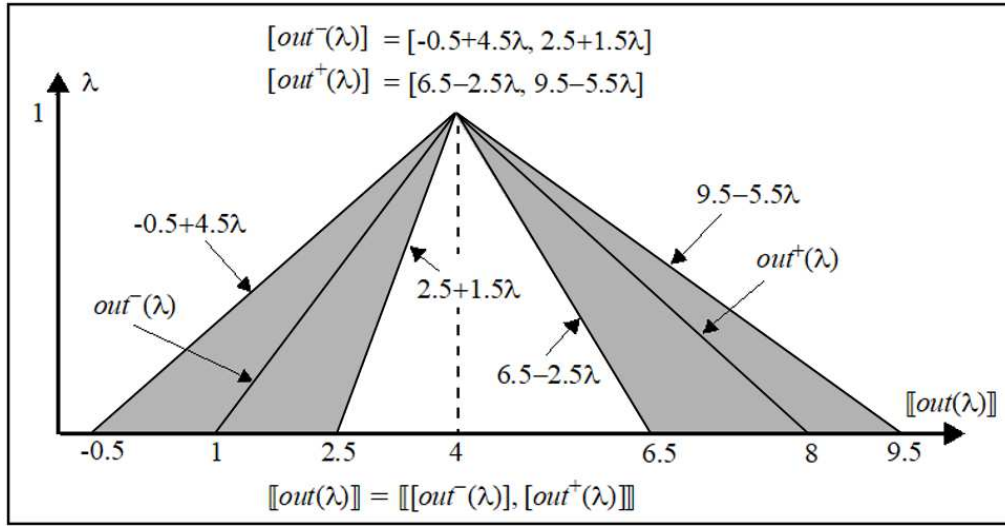
Therefore, these inputs are IV-T2FSs and can be represented by the following TGIs:

$$\begin{aligned} \llbracket in_1(\lambda) \rrbracket &= \llbracket [-1+3\lambda, 1+\lambda], [4-2\lambda, 6-4\lambda] \rrbracket \\ \llbracket in_2(\lambda) \rrbracket &= \llbracket [0.5+1.5\lambda, 1.5+0.5\lambda], [2.5-0.5\lambda, 3.5-1.5\lambda] \rrbracket \end{aligned}$$

Furthermore, the model output is also a TGI (IV-T2FS), which is obtained through an addition operation between TGIs (IV-T2FSs), i.e.,

$$\llbracket out(\lambda) \rrbracket = \llbracket in_1(\lambda) \rrbracket + \llbracket in_2(\lambda) \rrbracket = \llbracket [-0.5+4.5\lambda, 2.5+1.5\lambda], [6.5-2.5\lambda, 9.5-5.5\lambda] \rrbracket \quad (7)$$

By analogy with IV-T1FR, IV-T2FR is implemented through SIA over TGIs (see Appendix B).

Fig 8: IV-T2FSs (TGIs) $[[in_1(\lambda)]]$ and $[[in_2(\lambda)]]$ Fig 9: IV-T2FS (TGI) $[[out(\lambda)]]$

As illustrated in Figs. 8-9, the IV-T1FSs $[in_1(\lambda)]$ and $[in_2(\lambda)]$ are possible realizations of in_1 and in_2 . In all situations, if $[in_1(\lambda)] \in [[in_1(\lambda)]]$ and $[in_2(\lambda)] \in [[in_2(\lambda)]]$, then $[out(\lambda)] \in [[out(\lambda)]]$. The TGIs $[[in_1(\lambda)]]$, $[[in_2(\lambda)]]$ and $[[out(\lambda)]]$ can be viewed as a family of possibility distributions of the variables in_1 , in_2 and out , respectively.

4. Interval-valued fuzzy regression typology

For simplicity of interpretation and without loss of generality, a univariate context with crisp inputs and CI outputs is adopted in this section. Consider a set of N data $(x_i, [y_i])$, $i = 1, \dots, N$, which represents the inputs and outputs of a system. Let us assume the availability of a model \mathcal{M} that is able to fit the data, i.e.,

$$[\hat{y}] = \mathcal{M}(x, [\theta]); [\theta] = ([\theta_0], \dots, [\theta_M])^T; [\theta_j] = [\theta_j^-, \theta_j^+]; j = 0, \dots, M-1 \quad (8)$$

where x is the crisp input, $[\hat{y}]$ is the CI estimation produced by \mathcal{M} , $[\theta] \in \mathbb{I}^M$ ($M < N$) is the vector of CI parameters to be identified and \mathcal{M} is the input-output relationship, which can be considered linear, nonlinear, parametric or nonparametric. As discussed previously, CIR analysis is a non-SR method and is not based on probability theory. For this reason, CIR models do not have error terms; instead, they are contained in the CI coefficients. In the remainder of this section, all the analyzed concepts are illustrated on CIs, but they remain directly transposable on IV-T1FSs and IV-T2FSs. For instance, model (8) can be extended to IV-T1FSs and given by:

$$[\hat{y}(\lambda)] = \mathcal{M}(x, [\theta(\lambda)]); \text{ with: } [\theta(\lambda)] = ([\theta_0(\lambda)], \dots, [\theta_M(\lambda)])^T; [\theta_j(\lambda)] = [\theta_j^-(\lambda), \theta_j^+(\lambda)]; j = 0, \dots, M-1 \quad (9)$$

where $[\hat{y}(\lambda)]$ is the IV-T1FS output and $[\theta(\lambda)]$ is the vector of IV-T1FS parameters. In the same way, when the outputs are IV-T2FSs, model (9) can be expressed as follows:

$$[\hat{y}(\lambda)] = \mathcal{M}(x, [\theta(\lambda)]); [\theta(\lambda)] = ([\theta_0(\lambda)], \dots, [\theta_M(\lambda)])^T \quad (10)$$

where $[\theta_j(\lambda)] = [[\theta_j^-(\lambda)], [\theta_j^+(\lambda)]]; j = 0, \dots, M-1$, $[y(\lambda)]$ is the IV-T2FS output, and $[\theta(\lambda)]$ is the vector of IV-T2FSs (TGIS) parameters.

4.1. CIR: possibilistic and least squares approaches

The objective of this section is to analyze the philosophy of possibilistic and LS CIR methodologies through ontic and epistemic interpretations.

4.1.1. Least squares CIR

This approach is based on a scalar distance between set-valued data [5][30][34]. The CIR problem is to find the best fit model \mathcal{M} by minimizing the distance between observed and predicted CIs, typically a sum of squares of differences between upper and lower bounds of the observed outputs $[y_i]$ (at x_i) and the outputs $[\hat{y}_i]$ (at x_i) predicted by the model. This CIR corresponds to an optimization problem, i.e.,

$$\min_{[\theta]} \sum_{i=1}^N D^2([y_i], [\hat{y}_i]) = \min_{[\theta]} \sum_{i=1}^N D^2([y_i], \mathcal{M}(x_i, [\theta])) \quad (11)$$

where D is the Euclidean distance between CIs. Therefore, the founding idea of this CIR is based on the computation of distances between CIs. Considering two CIs $[a] = [a^-, a^+]$ and $[b] = [b^-, b^+]$, the Euclidean distance between them is given by:

$$D([a], [b]) = \sqrt{(a^- - b^-)^2 + (a^+ - b^+)^2} \quad (12)$$

For example, if $[a] = [1, 6]$ and $[b] = [4, 10]$, then $D = 5$. Furthermore, definition (12) is inspired by the representation of complex numbers and assumes that the CI endpoints can be seen as a point in \mathbb{R}^2 according to the EP diagram [59] (see Fig. 10, where lb is the lower bound and ub is the upper bound).

The evaluation of the distance D with a single and crisp scalar is consistent with the ontic approach where a CI is considered as a compact entity and can be characterized by a point in the EP diagram (the distance between intervals is regarded as the distance between their corresponding 2D points in the EP diagram). In Fig. 10, improper CIs are intervals with negative radii. However, a problem arises immediately when the CI is epistemic. Therefore, the CI endpoints represent a nonsingleton set, whereas a complex number is a point in \mathbb{R}^2 (a singleton). Unlike ontic CIs where the distance is a scalar, the distance between epistemic CIs should be an epistemic CI. This finding is in concordance with the legitimate question in [19]: “if we do not know the numbers exactly, how can the distance between them be an exact value?”. Moreover, in [19], the authors state that the distance between two numbers with uncertainty must be a number with uncertainty.

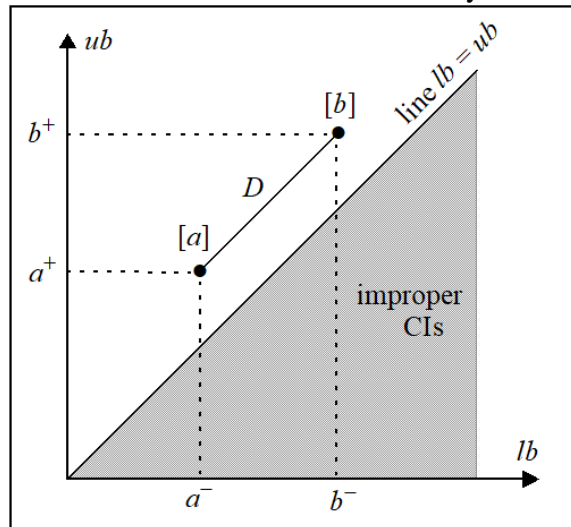


Fig. 10: Representation of the Euclidean distance between CIs in the EP diagram

From philosophical perspectives, this reasoning is coherent and is best suited for epistemic CIs. Furthermore, in [76], a generalization of the Euclidean distance to CIs is proposed where the distance between two CIs is also a CI, without losing the characteristics of the Euclidean metric when it leads with real numbers. This distance, which is well adapted for epistemic CIs, is given by:

$$[D_e([a],[b])]=[\inf\{|a-b|: a\in[a] \text{ and } b\in[b]\}, \sup\{|a-b|: a\in[a] \text{ and } b\in[b]\}] \quad (13)$$

Therefore, the distance between $a\in[a]$ and $b\in[b]$ is $D_e(a, b)\in[D_e]$. A formulation of the distance (13) is proposed in [76] and given by:

$$[D_e([a],[b])]=\begin{cases} [a^- - b^+, a^+ - b^-]; & \text{if: } [b] \leq [a] \text{ and } [a] \cap [b] = \emptyset \\ [0, a^+ - b^-]; & \text{if: } [b] \leq [a] \text{ and } [a] \cap [b] \neq \emptyset \\ [0, \max(b^+ - a^-, a^+ - b^-)]; & \text{if: } [b] \subseteq [a] \end{cases} \quad (14)$$

An analysis of (13) makes it possible to affirm that it is a generalization of the Euclidean distance to epistemic CIs, i.e.,

$$[D_e([a],[b])]=\sqrt{([a]-[b])^2} = |[a]-[b]| = [a^- - b^+, a^+ - b^-] \quad (15)$$

where the absolute value of a CI $[x]$ is expressed as follows:

$$|[x]| = |[x^-, x^+]| = \begin{cases} [\min(|x^-|, |x^+|), \max(|x^-|, |x^+|)]; & \text{if } x^- x^+ \geq 0 \\ [0, \max(|x^-|, |x^+|)]; & \text{if } x^- x^+ < 0 \end{cases} \quad (16)$$

For example, if we consider two epistemic CIs $[a] = [1, 6]$ and $[b] = [4, 10]$, then $[a]-[b] = [-9, 2]$ and:

$$[D_e([1,6],[4,10])]=\sqrt{([-9,2])^2} = |[-9,2]| = [0,9]$$

Therefore, if we take $a = 2\in[a] = [1, 6]$ and $b = 9\in[b] = [4, 10]$, then $D_e = 7 > D = 5$. Moreover, if $a = 1.5\in[a]$ and $b = 5\in[b]$, then $D_e = 3.5 < D$. In the same way, if $a = 5\in[a]$ and $b = 5\in[b]$, then $D_e = 0$. These results are in opposition with the unique scalar distance between CIs given by (12).

To summarize, the practitioners of CIR must incorporate the fact that the least squares approach based on Euclidean distance considers the CIs data as ontic entities. In this framework, when the data are represented by epistemic CIs, although nothing forbids the use of this distance, special attention must be given to the applicability and the results inherent to this approach to avoid misinterpretations. Furthermore, and as explained in [25], the scalar distance between epistemic CIs can nevertheless be sometimes useful and can be viewed as a kind of informational distance between pieces of information, whose role can be similar to relative entropy for probability distributions.

4.1.2. Possibilistic CIR

The possibilistic CIR aims at determining the model parameters so that the observed data are included in the outputs that are predicted by the model [12][18][73]. This approach is not based on the minimization of a distance as for the least squares approach but on the minimization of the entire imprecision of the model by minimizing the total spread (radius) of the model output, subject to constraints that the support of the model estimations $[\hat{y}_i]$ covers the support of the observations $[y_i]$. This possibilistic CIR corresponds to the following constrained optimization problem:

$$\begin{cases} \min_{[\theta]} \sum_{i=1}^N R([\hat{y}_i]) = \min_{[\theta]} \sum_{i=1}^N R(\mathcal{M}(x_i, [\theta])) \\ \text{UC: } [y_i] \subseteq [\hat{y}_i]; i = 1, \dots, N; R[\theta_j] \geq 0, j = 0, \dots, M-1 \end{cases} \quad (17)$$

The radius positivity of CI parameters aims to obtain well-defined CIs. In the situation where the model \mathcal{M} depends linearly on $[\theta]$, the radius of the model outputs can be expressed by the radius of the parameters. In this framework, the method minimizes the entire uncertainty of the model by minimizing the total spread (radius) of the parameters. Therefore, the optimization problem can be implemented as a mathematical linear programming problem [9][24]. In situations where some constraints and/or optimization criteria are

nonlinear, nonlinear programming methodologies [21][48][60][70], especially quadratic programming, have been employed [21][60]. Since the pioneering work of Tanaka, this possibilistic method has been improved by several authors and more efficient versions of the optimization criterion have been proposed (see [24] for a review). In this paper, we limit ourselves to the simple criterion given in (17). Our goal is not to compare these criteria but to clarify the philosophical and methodological principles of the possibilistic CIR.

According to (17), it can be observed that the spread (radius) of a CI is the same regardless of the semantics associated with this CI: ontic or epistemic. Therefore, according to its philosophy, the possibilistic CIR can be applied to address ontic and/or epistemic CIs using the same formalism. This property can represent an advantage over the least squares method, which has allowed its development in different areas of applications. Thus, if the CIs are ontic, the obtained model exhibits an interval-valued (set-valued) phenomenon. In contrast, when the CIs are epistemic, it attempts to cover both the evolution and the tendency of the model and the evolution of the amount of knowledge of this phenomenon [12][25]. In the remainder of this paper, the possibilistic approach is adopted. However, the least squares approach can be considered in a similar manner while taking into account its specificities. It is important to emphasize that the possibilistic approach always leads to a CI model, even in the presence of crisp data [12].

4.2. CIR: linear/nonlinear, parametric/nonparametric

4.2.1. Linear and nonlinear parametric models

Generally, solving the regression problem by introducing more parameters into the model can lead to overfitting. However, choosing too few parameters (a simplistic model) can result in bad fits. Thus, finding a good compromise for the number of parameters is part of the challenge of building good models.

A parametric model (PM) assumes that its vector of parameters $[\theta]$ is able to capture all the phenomena present in the data. Therefore, the model complexity and the associated number of parameters are limited (fixed) even if the amount of data is not limited. This property can limit the flexibility of these PMs. In CIR, most of the research on regression has focused on some predefined parametric forms, especially on linear models. For instance, in a linear framework, the univariate interval model (8) becomes:

$$[\hat{y}] = \mathcal{M}(x, [\theta]) = [\theta_0] + [\theta_1] \times x; [\theta] = ([\theta_0], [\theta_1])^T \quad (18)$$

As mentioned previously, all CI models can be naturally extended to IV-T1FS and IV-T2FS models. For example, models (9) and (10) can be reformulated in IV-T1FS and IV-T2FS contexts as follows:

$$[\hat{y}(\lambda)] = \mathcal{M}(x, [\theta(\lambda)]) = [\theta_0(\lambda)] + [\theta_1(\lambda)] \times x: \text{ for an IV-T1FS (GI) model} \quad (19)$$

$$[\hat{y}(\lambda)] = \mathcal{M}(x, [\theta(\lambda)]) = [\theta_0(\lambda)] + [\theta_1(\lambda)] \times x: \text{ for an IV-T2FS (TGI) model} \quad (20)$$

The linear models sometimes remain limited in the presence of complex data. Therefore, they are, however, not suitable for all circumstances. Considering the complexity and potential nonlinearity of real-world data, nonlinear models can be envisioned. A crucial point is then the choice of the nonlinear model shape. For their simplicity of exploitation and their adaptability with the optimization methodologies initially proposed for linear regression, some well-known models, namely, polynomial and piecewise multilinear models, are often used. These models are expressed as follows:

$$[\hat{y}] = \mathcal{M}(x, [\theta]) = [\theta_0] + \sum_{j=1}^M [\theta_j] \times x^j : \text{ for a polynomial CI model} \quad (21)$$

$$[\hat{y}] = \mathcal{M}(x, [\theta]) = \bigcup_{k=1}^{k=N_c} ([\theta_0^k] + [\theta_1^k] \times x) : \text{ for a piecewise linear CI model} \quad (22)$$

In practical scenarios, the piecewise linear approach remains sometimes difficult to apply, where particular attention must be given to the problems of data splitting and the junction between the N_c piecewise models [9][12][81]. The polynomial approach may be effective in some cases but may lead to a problem of oscillation of higher-order polynomial models (the well-known Runge's phenomenon). These approaches have been analyzed in [9][10][12] with application examples. Although these models are nonlinear, they

have the particularity to depend linearly on $[\theta]$. However, in general, for nonlinear regression, the model \mathcal{M} cannot be written as linear with regard to $[\theta]$. Exponential (or logarithm) models are classic examples of this category. These nonlinear models are useful in some fields, such as ecology, agriculture, and biology. For example, the following exponential model is nonlinear, where the parameters are nonlinearly dependent on the parameters [42]:

$$\mathcal{M}(x, [\theta]) = [\theta_0] \times (1 - e^{[\theta_1] \times x}); [\theta] = ([\theta_0], [\theta_1])^T \quad (23)$$

Some nonlinear regression functions can be linearized through input-output transformation [42]. In the abovementioned linear and/or nonlinear PMs, the IO mapping \mathcal{M} is assumed to be known. However, in many situations, the relationship \mathcal{M} is not known. In this context, the philosophy of nonparametric models (NPMs), which are generally nonlinear, can be exploited.

4.2.2. Nonparametric models

NPMs differ from PMs in the sense that the shape of \mathcal{M} is not predetermined but can be adjusted to capture unexpected features of the data. For example, K-nearest neighbors, RBF kernel, SVMs, local polynomials, splines, wavelets, etc., (e.g., [46][58][62][77][83]) are regarded as NPMs since the number of parameters grows with the size of the training data.

Unlike PMs, NPMs assume that the data cannot be represented by a set of finite parameters, i.e., they assume an infinite dimension of $[\theta]$. The quantity of information that $[\theta]$ can capture increases as the quantity of the data increases. This property makes NPMs more flexible than PMs.

Despite their lack of performance and flexibility in some complex situations, PMs have advantages of interpretability, where their parameters have some physical meaning, which makes them interpretable and of interest. Moreover, PMs offer some advantages of transferability and theoretical understanding that are lacking in nonparametric models. From a lexical point of view, the term nonparametric does not mean that such models have no parameters but that the number of parameters is flexible and not fixed in advance. Notably, even NPMs are finally parameterized. Unlike a PM that uses a fixed number of parameters, each of them having a fixed meaning, an NPM has an indefinite number of parameters, which depends on the data. On a philosophical level, there is inherently no fundamental difference between PMs and NPMs. Any mapping can be perfectly approximated by a model with an infinite number of parameters, but yet parametric. In our opinion, the trade-off between the two parametric and nonparametric paradigms, especially in an interval framework, is in computational cost and accuracy. A PM is computationally lighter but requires stronger assumptions about the data. In contrast, an NPM is computationally heavier but requires fewer assumptions about the data. Generally, NPMs are more efficient for modeling complex nonlinear data [49]. In a simple way, when the IO mapping is known, PMs should be used. If the relationship is unknown (no prior knowledge), NPMs should be envisioned. To summarize, the aim of the nonparametric regression is to identify the best regression function according to the data distribution rather than estimating the parameters of a specific model. As an example of NPMs that we will use in this paper is the B-splines formalism [26][27][56][78]. Thus, due to their properties, such as local shape controllability, boundedness, continuity, etc., the B-splines are good candidates for fitting complex data. In a CIR framework, the model \mathcal{M} is represented by an interval B-spline function as follows:

$$[\hat{y}] = \mathcal{M}(x, [\theta]) = \sum_{j=0}^{M-1} [\theta_j] \cdot B_{j,k}(x); x \in [t_0, t_{n+k-1}] \quad (24)$$

where $[\theta_j]$ are CIs representing the control coefficients to be identified and $B_{j,k}(x)$ are the B-spline basis functions [26][27] defined on a sequence of knots. In (24), the values $x = t_j, j = 0, \dots, M+k-1$ are the knots, where M represents the number of control coefficients and k denotes the B-spline order. The B-spline of order k is a piecewise polynomial function of degree $k-1$. Considering x in the domain $[x_0, x_f]$, the following nondecreasing sequence represents the knot sequence:

$$T = \{x_0 = t_0 = \dots = t_{k-1}, t_k, \dots, t_{M-1}, t_M = \dots = t_{M+k-1} = x_f\} \quad (25)$$

According to (25), it can be stated that the first and last knots are of multiplicity k . This condition is assumed for obtaining clamped B-splines. Model (24) is a linear combination of the basic functions $B_{j,k}(x)$

defined on the knot sequence T with M control coefficients $[\theta_j(\lambda)]$; $j = 0, \dots, M-1$ [56][78]. As for PMs, the B-spline approach can be naturally extended to IV-T1FS and IV-T2F NPMs. In the IV-T1FS context, model (24) becomes:

$$[\hat{y}(\lambda)] = \mathcal{M}(x, [\theta(\lambda)]) = \sum_{j=0}^{M-1} [\theta_j(\lambda)] \cdot B_{j,k}(x) \quad (26)$$

where $[\theta_j(\lambda)]$, $j = 0, \dots, M-1$ are the IV-T1FS control coefficients to be identified.

In the same way, the IV-T2FS B-spline model is given by:

$$[\hat{y}(\lambda)] = \mathcal{M}(x, [\theta(\lambda)]) = \sum_{j=0}^{M-1} [\theta_j(\lambda)] \cdot B_{j,k}(x) \quad (27)$$

In (27), $[\theta_j(\lambda)]$, $j = 0, \dots, M-1$ represent the IV-T2FS control coefficients to be identified and $[\hat{y}(\lambda)]$ is the IV-T2FS model output.

5. Interval-valued fuzzy possibilistic regression methodology

In the literature, the regression problem with fuzzy data has been addressed from different points of view and successfully applied in different applications. As discussed in the paper introduction, IV-T1FR aroused a major interest. However, IV-T2FR remains a little-studied field where no generic methodology for its implementation is proposed. The originality of this work consists of using a generic and a unified approach based on the use of GIs and TGIs for carrying out IV-T1FR and IV-T2FR.

In this section, the possibilistic approach is adopted to represent the input-output behavior. It is well known that this approach is sensitive to outliers, and a filtering procedure is necessary before its implementation. For simplicity and without loss of generality, this section considers a subproblem of the general regression problem where the model \mathcal{M} depends linearly on the parameter vector.

5.1. Possibilistic IV-T1FR methodology

The philosophy of the IV-T1FR finds its founding principles and justifications through the following situations:

- The instantiation of measures in their likelihood (membership) CIs, to which degrees of confidence and/or certainty are associated.
- The manipulation of subjective information is represented by distributions of possibility (IV-T1FSs) that are defined by experts.

When considering a set of IO data $\{(x_i, [y_i(\lambda)])\}$, $i = 1, \dots, N$, where x_i is the i^{th} input and $[y_i(\lambda)]$ is the corresponding IV-T1FS output at x_i , the IV-T1FR objective is to find an IV-T1FS model of the form (9). This IV-T1FR aims at determining the IV-T1FS parameter vector $[\theta(\lambda)]$ such that the observed outputs $[y_i(\lambda)]$ are included in the outputs generated by the model, i.e., $[y_i(\lambda)] \subseteq [\hat{y}_i(\lambda)]$. This IV-T1FR is formalized by the following optimization problem under constraints (UC):

$$\begin{aligned} \min_{[\theta(\lambda)]} \sum_{i=1}^N R([\hat{y}_i(\lambda)]) &= \min_{[\theta(\lambda)]} \sum_{i=1}^N R(\mathcal{M}(x_i, [\theta(\lambda)])) \\ \text{UC: } [y_i(\lambda)] &\subseteq [\hat{y}_i(\lambda)]; \quad i = 1, \dots, N; R([\theta_j(\lambda)]) \geq 0, \quad j = 0, \dots, M-1 \end{aligned} \quad (28)$$

The inclusion of two CIs $[a] = [a^-, a^+]$ and $[b] = [b^-, b^+]$ is defined by $[a] \subseteq [b] \Leftrightarrow b^- \leq a^- \text{ \& } a^+ \leq b^+$

In (28), $R([\hat{y}_i(\lambda)])$ represents the radius of the model output. The constraints are imposed to ensure that all observed data $[y_i(\lambda)]$ are included in the predicted outputs $[\hat{y}_i(\lambda)]$. In the optimization problem (28), the model vagueness is represented by the sum of the radii of the outputs. The radius positivity constraints of $[\theta_j(\lambda)]$, $j = 0, \dots, M-1$ must also be taken into account to obtain well-defined IV-T1FSs.

In practical implementations, the optimization problem (28) is realized by an approximation using the CI approach by discretizing the λ -vertical dimension. In this context, a compromise between the computation time and the quality of the approximation must be achieved. In common situations, which are often exploited in the literature, where the observed outputs are triangular or trapezoidal IV-T1FSs, the optimization procedure (28) can be limited to the levels $\lambda = 0$ and $\lambda = 1$, i.e.,

$$\begin{aligned}
& \min_{[\theta(0)], [\theta(1)]} \sum_{i=1}^N R([\hat{y}_i(0)]) + R([\hat{y}_i(1)]) = \min_{[\theta(0)], [\theta(1)]} \sum_{i=1}^N R(\mathcal{M}(x_i, [\theta(0)])) + R(\mathcal{M}(x_i, [\theta(1)])) \\
& \text{UC: } [\hat{y}_i(0)] \subseteq [y_i(0)]; [\hat{y}_i(1)] \subseteq [y_i(1)]; i = 1, \dots, N \\
& R([\theta_j(0)]) \geq 0, R([\theta_j(1)]) \geq 0, j = 0, \dots, M-1 \\
& \hat{y}_i(1) \subseteq \hat{y}_i(0), i = 1, \dots, N
\end{aligned} \tag{29}$$

In (29), additional inclusion constraints for the level $\lambda = 1$ in the level $\lambda = 0$ are necessary to guarantee well-defined IV-T1FSs. Therefore, if the model expression \mathcal{M} is known and/or linearly dependent on the parameters, the optimization problems (28) and (29) are simplified. This point is discussed in section 6.

5.2. Possibilistic IV-T2FR methodology

The philosophy of the IV-T2FR methodology finds its founding principles and its justifications in the following situations:

- When the IV-T1FR is insufficient to exhibit the presence of uncertainty in the IV-T1FS representation. Therefore, the locations of IV-T1FS (GI) profiles are not precisely known or are hidden for confidentiality purposes.
- When the integration of uncertainty into the answers of experts and the simultaneous consideration of their different opinions are needed.
- When the input-output data are stated in terms of IV-T2FSs.

Let us consider a set of IO data $\{x_i, [y_i(\lambda)]\}; i = 1, \dots, N$. By analogy with the IV-T1FR, the IV-T2FR aims to determine an IV-T2FS (TGI) model in the form (10). The IV-T2FR objective is to determine the IV-T2FS (TGI) parameter vector $[\theta(\lambda)] = ([\theta_0(\lambda)], \dots, [\theta_M(\lambda)])^T$ such that the observed outputs $[y_i(\lambda)]$ are included in the IV-T2FSs (TGIs) produced by the model, i.e., $[y_i(\lambda)] \subseteq [\hat{y}_i(\lambda)]$. This IV-T2FR problem can be formalized by the following optimization problem under constraints (UC):

$$\begin{aligned}
& \min_{[\theta(\lambda)]} \sum_{i=1}^N R([\hat{y}_i(\lambda)]) = \min_{[\theta(\lambda)]} \sum_{i=1}^N R(\mathcal{M}(x_i, [\theta(\lambda)])); \\
& \text{UC: } [y_i(\lambda)] \subseteq [\hat{y}_i(\lambda)]; i = 1, \dots, N \\
& R([\theta_j^-(\lambda)]) \geq 0, R([\theta_j^+(\lambda)]); j = 0, \dots, M-1 \\
& [\theta_j^-(\lambda)] \leq [\theta_j^+(\lambda)]; j = 0, \dots, M-1;
\end{aligned} \tag{30}$$

In (30), $R([\hat{y}_i(\lambda)]) = R([\hat{y}_i^-(\lambda)]) + R([\hat{y}_i^+(\lambda)])$. Furthermore, the inclusion relation between two TIs $[a] = [[a^-], [a^+]]$ and $[b] = [[b^-], [b^+]]$ is defined as follows:

$$[a] \subseteq [b] \Leftrightarrow ([b^-] - [a^-] \in \mathfrak{R}^-) \& ([a^+] - [b^+] \in \mathfrak{R}^-) \tag{32}$$

The constraint $[\theta_j^-(\lambda)] \leq [\theta_j^+(\lambda)]$ in (30) guarantees that $[\theta_j^-(\lambda)]$ is always before $[\theta_j^+(\lambda)]$ with no intersection between them for $\lambda \in [0, 1]$ and $[\theta_j^-(1)]$ is always before $[\theta_j^+(1)]$ with a possible meeting between them. Moreover, for obtaining well-defined parameters, the positivity of the radii of the left and right profiles is also required, i.e., $R([\theta_j^-(\lambda)]) \geq 0$ and $R([\theta_j^+(\lambda)]) \geq 0$. In some practical applications, the observed outputs are given by uncertain IV-T1FSs instead of IV-T2FSs. Therefore, due to the presence of uncertainty in the IV-T1FS observations, for each input x_i , a number P of IV-T1FSs can be observed (P process repetitions). Therefore, the observed data are now $\{x_i, [y_{i,p}(\lambda)]\}; i = 1, \dots, N, p = 1, \dots, P$. In this context, the inclusion constraint $[y_i(\lambda)] \subseteq [\hat{y}_i(\lambda)]$, $i = 1, \dots, N$ in (30) is replaced by the constraint $[y_{i,p}(\lambda)] \subseteq [\hat{y}_i(\lambda)]$, $i = 1, \dots, N, p = 1, \dots, P$. As for the IV-T1FR where the observed outputs have trapezoidal shapes, the optimization can be limited to the levels $\lambda = 0$ and $\lambda = 1$. Furthermore, if the model expression \mathcal{M} is known and/or linearly dependent on the parameters, the optimization problem (30) can be simplified (see the next section).

6. Illustrative examples

This section aims to show how the IV-T1FR and the IV-T2FR regression models are constructed using the proposed unified methodology. Therefore, this section does not focus on showing the superiority of the

proposed method in terms of numerical performances but rather suggests a unifying framework that applies to whatever the optimization criterion and the model used. In this paper, the simple and well-known criterion that corresponds to the sum of the output radii has been used. Furthermore, all the interval-valued type-1 and type-2 fuzzy regression methods that are published in the literature can be approached with the proposed formalism. For comparison purposes, two benchmark examples are added in Appendix C.

The implementation of the IV-T1FR and the IV-T2FR through PMs and NPs are illustrated using a univariate data set given in [46]. The inputs are crisp, and the outputs are symmetrical triangular IV-T1FSs. The data set $\{(x_i, [y_i(\lambda)])\}, i = 1, \dots, 51$, where x_i is the i^{th} input and $[y_i(\lambda)] = [y_i^o - (1-\lambda)e_i, y_i^o + (1-\lambda)e_i]$ is the corresponding triangular IV-T1FS output at x_i . These IO data are generated by:

$$\begin{cases} x_i = 0.02(i-1); & y_i^o = (2.7x_i - 0.2)^2 + 4.5 + \text{err}_i \\ e_i = 1.7 \exp(-49(x_i - 0.5)^2) + 1.7x_i + 1.2 \end{cases} \quad (31)$$

The noise err_i in (31) has been obtained using a uniform distribution over the interval $[-0.4, 0.4]$.

6.1. IV-T1FS based regression

6.1.1. Parametric approach

Let us assume that the PM is linear and is given by (19). The IV-T1FR objective is to determine the IV-T1FS parameters $[\theta_0(\lambda)]$ and $[\theta_1(\lambda)]$. In this example, a trapezoidal model where the IV-T1FSs have a trapezoidal shape is employed to ensure the satisfaction of the inclusion constraints (see [9] for the justification of the trapezoidal model). Due to the trapezoidal shapes of the parameters, the optimization problem (29) is applied. Furthermore, the radii of the model at the levels $\lambda=0$ and $\lambda=1$ are given by:

$$\begin{cases} R(\mathcal{M}(x_i, [\theta(0)])) = R([\theta_0(0)]) + R([\theta_1(0)]) \times x_i; & x_i \geq 0 \\ R(\mathcal{M}(x_i, [\theta(1)])) = R([\theta_0(1)]) + R([\theta_1(1)]) \times x_i \end{cases}$$

the optimization procedure (29) is reduced to:

$$\begin{aligned} & \min_{[\theta_j(0)], [\theta_j(1)]} \sum_{i=1}^{51} R([\theta_0(0)]) + R([\theta_1(0)]) + (R([\theta_0(1)]) + R([\theta_1(1)]) \times x_i) \\ & \text{UC: } [\hat{y}_i(0)] \subseteq [y_i(0)]; [\hat{y}_i(1)] \subseteq [y_i(1)]; i = 1, \dots, 51 \\ & R([\theta_j(0)]) \geq 0, R([\theta_j(1)]) \geq 0, j = 0, 1 \\ & \hat{y}_i(1) \subseteq \hat{y}_i(0), i = 1, \dots, 51 \end{aligned}$$

The implementation of this procedure gives the following IV-T1FS parameters:

$$[\theta_0(\lambda)] = [0.69 + 1.93\lambda, 6.21 - 1.65\lambda] \text{ and } [\theta_1(\lambda)] = [2.5 + 3.41\lambda, 7.36 - 1.45\lambda]$$

For example, when considering the parameter $[\theta_0(\lambda)]$, the optimization procedure leads to the CIs: $[\theta_0(0)] = [0.69, 6.21]$ (at $\lambda = 0$) and $[\theta_0(1)] = [2.62, 4.56]$ (at $\lambda = 1$). Using a linear interpolation between the levels 0 and 1, the IV-T1FS $[\theta_0(\lambda)]$ is obtained (see Fig. 11.b). The linear IV-T1FR led to the model illustrated in Fig. 11.b. According to the model parameters and thanks to the SIA operations over GIs, it is possible to express the model output at each input x . For example, the output at $x = 0.1$ is given by:

$$\begin{aligned} [\hat{y}_x(\lambda)] &= [\theta_0(\lambda)] + [\theta_1(\lambda)] \times x = [0.69 + 1.93\lambda, 6.21 - 1.65\lambda] + [2.5 + 3.41\lambda, 7.36 - 1.45\lambda] \times (0.1) \\ &= [0.94 + 2.27\lambda, 6.95 - 1.79\lambda] \end{aligned}$$

To measure the performance and the goodness of fit, the two indicators R^2 and $RMSE$ (root mean square error) defined in [23] are exploited. Due to the complexity of the data, it is clear from the R^2 and $RMSE$ indicators ($R^2 = 0.03$; $RMSE = 2.44$) that a linear model is insufficient to fit the data appropriately. To improve IV-T1FR performance, we can move toward a polynomial model. In this framework, the order of the polynomial model is primordial. This order can be tuned empirically. In this paper, this problem is addressed using a model selection approach according to the Akaike information criterion (AIC) [1]. This method can quantify the performance of a model and its complexity. Therefore, by comparing the models with different orders, it is possible to select the most suitable model. The application of the AIC method leads to an optimal order $n = 2$. In this case, the polynomial PM is given by:

$$[\hat{y}(\lambda)] = \mathcal{M}(x, [\theta(\lambda)]) = [\theta_0(\lambda)] + [\theta_1(\lambda)] \times x + [\theta_2(\lambda)] \times x^2$$

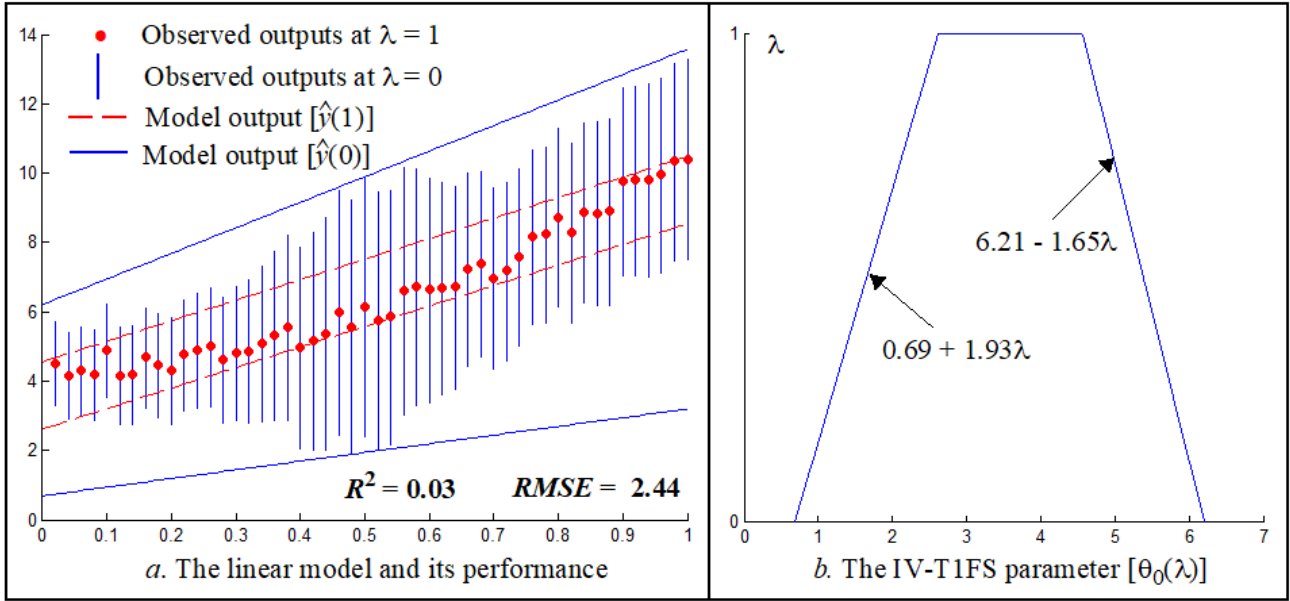


Fig. 11: Results of the IV-T1FR using a linear model

The application of the IV-T1FR method gives the following parameters:

$$[\theta_0(\lambda)] = [2.50 + 1.69\lambda, 7.49 - 2.56\lambda]; [\theta_1(\lambda)] = [-4.77 + 3.66\lambda, 1.16 - 2.23\lambda]; [\theta_2(\lambda)] = [7.27, 7.27]$$

The parameter $[\theta_2(\lambda)]$ is constant $\forall \lambda \in [0, 1]$. The polynomial model and its performance are shown in Fig. 12.a. Furthermore, for illustration, the parameter $[\theta_0(\lambda)]$ is given in Fig. 12.b.

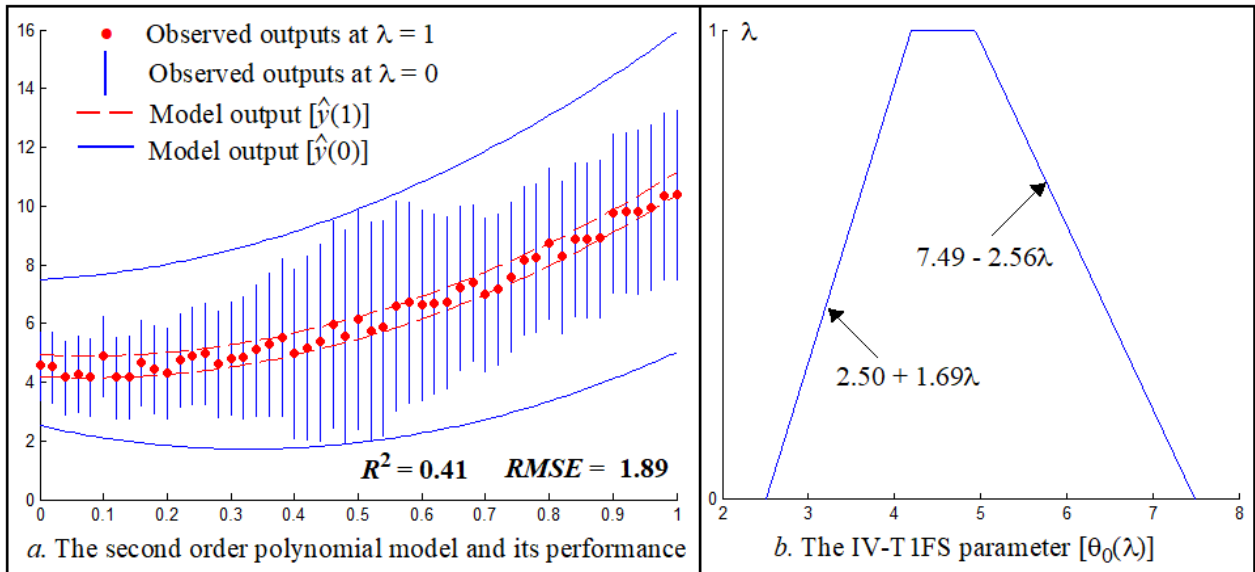


Fig. 12: Results of the IV-T1FR using a second-order polynomial model

Although the performances have been improved by the polynomial model ($R^2 = 0.41$; $RMSE = 1.89$), they remain insufficient. As the optimal model is obtained for $n = 2$, increasing the model order brings no significant improvement. However, to achieve a better fit, we can envisage other model forms (piecewise linear or other nonlinear forms) or move toward a nonparametric approach. Due to their undeniable performance in the presence of complex data, the solution of an NPM is preferred.

6.1.2. Nonparametric approach

To explore and fit the data more flexibly, the IV-T1FS B-spline model (26) is used. Other nonparametric models can be envisioned using the same methodology. In this paper, due to its performance, cubic splines ($k = 4$) are chosen. However, other B-splines with different degrees can also be used.

Knowing that the inputs are crisp and the control coefficients have trapezoidal shapes, the model outputs are also trapezoidal IV-T1FSs. Moreover, the trapezoidal shapes of the IV-T1FSs permit their expression only on levels $\lambda=0$ and $\lambda=1$. For illustration in the optimization procedure, the number M of control coefficients is chosen equal to 7. This choice is validated by the AIC as used in [47]. As the radii of model (26) at the levels, $\lambda=0$ and $\lambda=1$ are given by:

$$\begin{cases} R(\mathcal{M}(x_i, [\theta(0)])) = \sum_{j=0}^6 R([\theta_j(0)]) \cdot B_{j,4}(x) \\ R(\mathcal{M}(x_i, [\theta(1)])) = \sum_{j=0}^6 R([\theta_j(1)]) \cdot B_{j,4}(x) \end{cases}$$

The optimization problem (29) becomes:

$$\min_{[\theta_j(0)], [\theta_j(1)]} \sum_{i=1}^{51} \sum_{j=0}^6 (R([\theta_j(0)]) + R([\theta_j(1)]) \cdot B_{j,4}(x))$$

$$\text{UC: } [\hat{y}_i(0)] \subseteq [y_i(0)]; [\hat{y}_i(1)] \subseteq [y_i(1)]; i = 1, \dots, 51$$

$$R([\theta_j(0)]) \geq 0; R([\theta_j(1)]) \geq 0; j = 0, \dots, 6; \hat{y}_i(1) \subseteq \hat{y}_i(0); i = 1, \dots, 51$$

The IV-T1FR led to a B-spline model of the form (26), which is illustrated in Fig. 13.a. The control coefficients $[\theta_j(\lambda)]$ are IV-T1FSs. For example, the IV-T1FS $[\theta_2(\lambda)]$ is given by $[\theta_2(\lambda)] = [2.54 + 1.75\lambda, 7.10 - 1.97\lambda]$ and shown in Fig. 13.b. Therefore, the optimization gives the CIs: $[\theta_2(0)] = [2.54, 7.10]$ and $[\theta_2(1)] = [4.29, 5.13]$. By a linear interpolation between $\lambda = 0$ and $\lambda = 1$, $[\theta_2(\lambda)]$ is obtained.

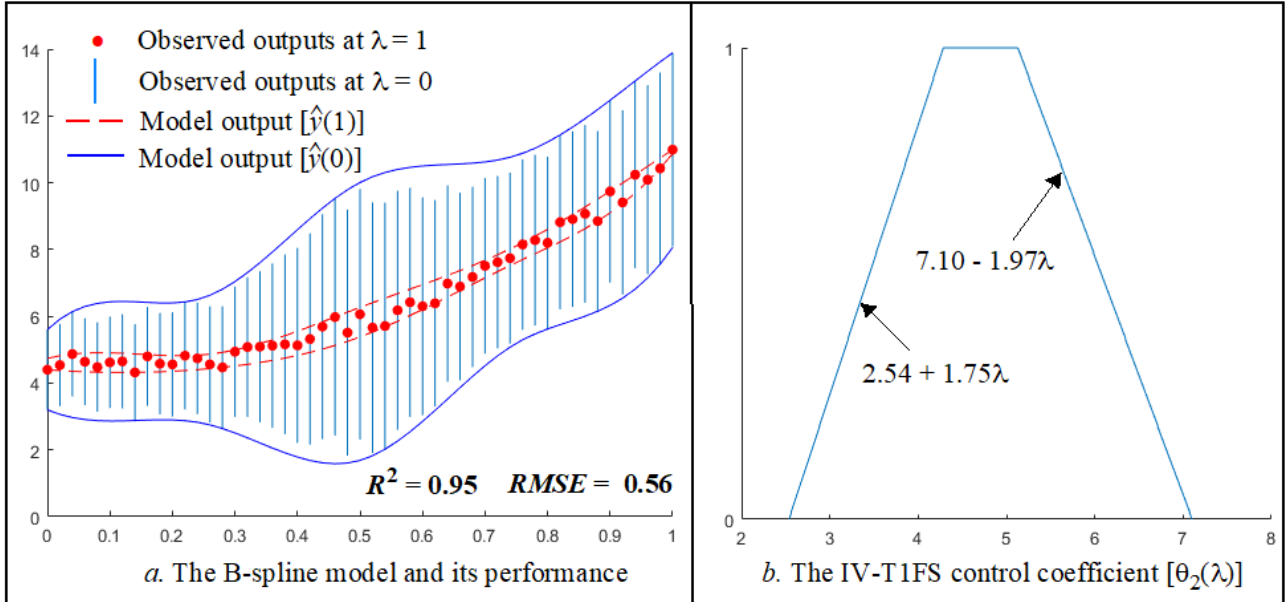


Fig. 13: Results of the IV-T1FR using a B-spline model

Compared to the linear and polynomial PMs, the better performance of the B-spline model is justified by the fact that it uses six third-order polynomials (cubic splines) instead of a single linear second-order polynomial model. In this illustration, the knot sequence is $T = \{0, 0, 0, 0, 1/6, 1/3, 1/2, 2/3, 5/6, 1, 1, 1, 1\}$, which advocates that the abscissa axis is divided into 6 regions: $[0, 1/6]$, $[1/6, 1/3]$, $[1/3, 1/2]$, $[1/2, 2/3]$, $[2/3, 5/6]$ and $[5/6, 1]$. In each region, a polynomial of degree 3 is used. Furthermore, the polynomials are constrained so that they join smoothly at the knots (at the region bounds). Table 2 shows the results of the proposed method using the vagueness criterion at the level $\lambda = 0$, i.e.,

$$\text{Vagueness} = (1/N) \cdot \sum_{i=1}^N R([\hat{y}_i(0)])$$

Used approach	Vagueness
The support vector machine (RBF kernel of Hao)	3.064
The IV-T1FR based on B-splines	2.817

Table 2: Comparative Results

The results of the IV-T1FR are comparable to those of Hao's method based on an SVM [46]. If both approaches can fit the fuzzy data effectively, they are not armed to cope with the uncertainty in the IV-T1FS representation. Therefore, if the process (data generated by (31)) is repeated P times, at each time, the IV-T1FR gives a different model with different IV-T1FS parameters. This phenomenon is induced by the uncertain nature of the used data. The objective is now to propose an IV-T2FR that can fit all the uncertain IV-T1FS behaviors in an IV-T2FS (TGI) model. Due to the presence of uncertainty, when the IV-T1FR process is repeated 15 times (at each process, the data are regenerated), the performance of the suggested approach is maintained (see Fig. 14, where the boxplots of the R^2 and the vagueness indicators are shown).

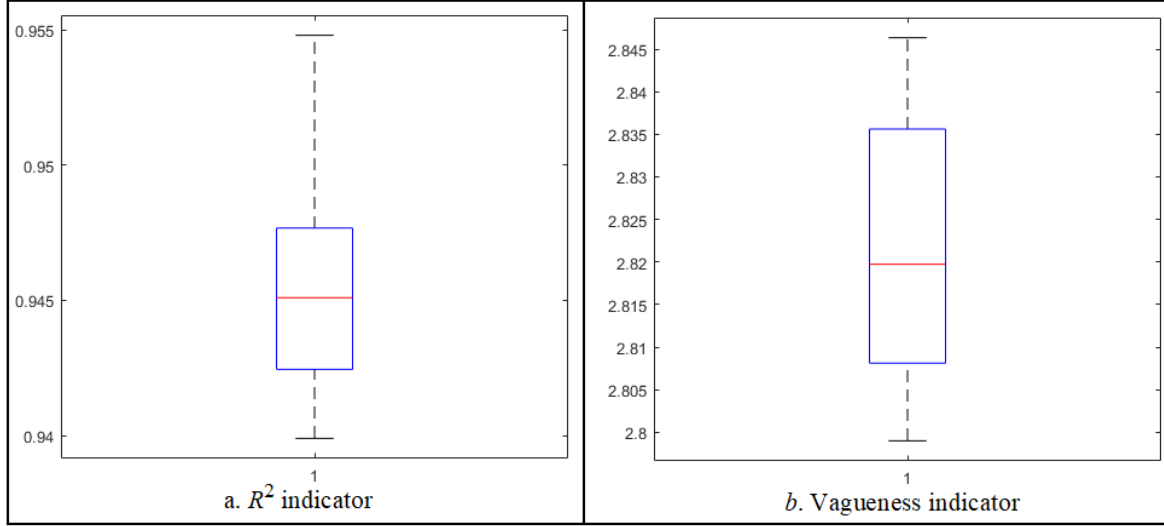


Fig. 14: Boxplots of the R^2 and the vagueness indicators

6.2. IV-T2FS based regression

As for the IV-T1FR, IV-T2FR can be envisioned through PMs and NPMs.

6.2.1. Linear parametric regression

Considering a set of observed trapezoidal IV-T1FSs that are obtained by model (19) through 15 process repetitions, $[y_{i,p}(\lambda)]$, $i = 1, \dots, 51$; $p = 1, \dots, 15$. The IV-T2FR aims at obtaining a linear model in the form (20), where $[\theta_0(\lambda)]$ and $[\theta_1(\lambda)]$ are TGIs (IV-T2FSs). Due to the trapezoidal shape of the model and its parameters, the optimization problem can be made only on the levels $\lambda = 0$ and $\lambda = 1$. In this case, as:

$$\begin{cases} R([\hat{y}_i(0)]) = R([\hat{y}_i^-(0)]) + R([\hat{y}_i^+(0)]) \\ R([\hat{y}_i(1)]) = R([\hat{y}_i^-(1)]) + R([\hat{y}_i^+(1)]) \end{cases}$$

The optimization problem (30) is implemented as follows:

$$\begin{aligned} \min_{[\theta_0(0)], [\theta_1(0)]} & \sum_{i=1}^{51} R([\hat{y}_i(0)]) + R([\hat{y}_i(1)]) \\ \text{UC: } [y_{i,p}(0)] \subseteq [\hat{y}_i(0)]; [y_{i,p}(1)] \subseteq [\hat{y}_i(1)]; & i = 1, \dots, 51; p = 1, \dots, 15 \\ R([\theta_j^-(0)]) \geq 0; R([\theta_j^+(0)]) \geq 0; R([\theta_j^-(1)]) \geq 0; & R([\theta_j^+(1)]); j = 0, 1 \\ [\theta_j^-(0)] \leq [\theta_j^+(0)]; [\theta_j^-(1)] \leq [\theta_j^+(1)]; & j = 0, 1; \end{aligned} \quad (32)$$

The application of the IV-T2FR method leads to the following IV-T2FS parameters:

$$\begin{aligned} [\theta_0(\lambda)] &= [[-1.26+3.56\lambda, 3.01+0.19\lambda], [5.74-1.23\lambda, 6.09-1.2\lambda]] \\ [\theta_1(\lambda)] &= [[1.69+3.67\lambda, 2.09+3.43\lambda], [7.76-1.8\lambda, 7.93-1.73\lambda]] \end{aligned}$$

For example, the IV-T2FS $[\theta_0(\lambda)]$ is built as follows: knowing that the optimization problem leads the TIs:

$$\begin{aligned} [\theta_0(0)] &= [\theta_0^-(0), \theta_0^+(0)] = [-1.26, 3.01], [5.74, 6.09]: \text{ at the level } \lambda = 0 \\ [\theta_0(1)] &= [\theta_0^-(1), \theta_0^+(1)] = [2.3, 3.20], [4.51, 4.89]: \text{ at the level } \lambda = 1 \end{aligned}$$

by linear interpolation between the levels $\lambda = 0$ and $\lambda = 1$, the TGI (IV-T2FS) coefficient $[\theta_0(\lambda)]$ shown in Fig. 15.b is obtained. In this context, for each input x , the IV-T2FS output (TGI) can be expressed analytically through the model (20). For instance, the output at $x = 0.1$ corresponds to the following expression (see Fig. 15.a):

$$[\hat{y}(\lambda)] = [\theta_0(\lambda)] + [\theta_1(\lambda)] \times (0.1) = [[-1.09 + 3.93\lambda, 3.22 + 0.53\lambda], [6.52 - 1.41\lambda, 6.88 - 1.37\lambda]]$$

It can be observed that all the uncertain IV-T1FSs are included in the IV-T2FS output (refer to Fig. 15.a).

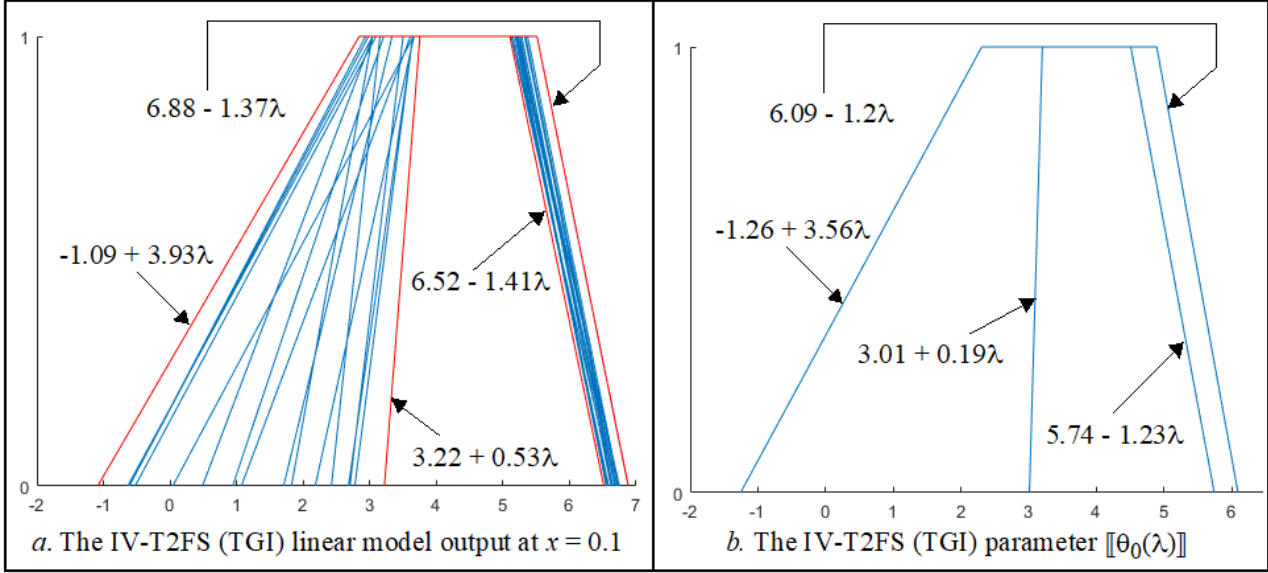


Fig. 15: Results of the IV-T2FR using a linear model

6.2.2. B-splines nonparametric regression

Let us consider the data set as presented in the previous subsection. When the B-spline model is always assumed, the IV-T2FR aims at obtaining a model in the form (27). The control coefficients are estimated according to the simplified optimization problem (32), where the model output is replaced by (27) with the corresponding number of IV-T2FSs $[\theta_i(\lambda)]$. The application of the IV-T2FR method leads to control coefficients that are IV-T2FSs (TGIs). For example, $[\theta_2(\lambda)]$ is illustrated in Fig. 16.b and given by:

$$[\theta_2(\lambda)] = [[2.20 + 1.71\lambda, 2.60 + 1.57\lambda], [5.61 - 1.32\lambda, 6.77 - 1.94\lambda]]$$

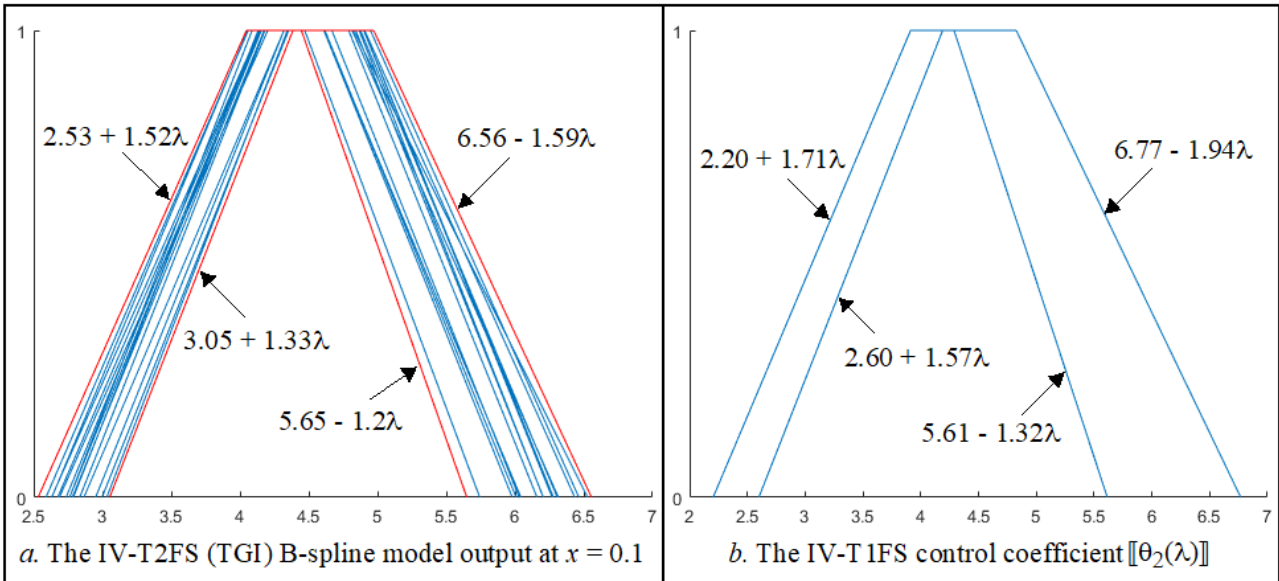


Fig. 16: Results of the IV-T2FR using a B-spline model

As for the IV-T1FR, at each input x , the output of the IV-T2FS model can be determined (see Fig. 16.a). For example, the output for $x = 0.1$ is given by:

$$[\hat{y}(\lambda)] = \sum_{j=0}^6 [\theta_j(\lambda) \cdot B_{j,4}(0.1)] = [[2.53+1.52\lambda, 3.05+1.33\lambda], [5.65-1.2\lambda, 6.56-1.59\lambda]]$$

As stated in Fig 16, the IV-T2FR model encapsulates all the IV-T1FR models. Our method has the advantage of being able to analytically express the model outputs. Furthermore, all the operations are implemented by an extension of the SIA to TGIs. Therefore, the rigor of SIA is conserved in IV-T2FR.

7. Remarks, limitations and perspectives

7.1. Optimization procedure

The fuzzy regression (IV-T1FR and IV-T2FR) is implemented as an optimization problem under constraints. Fig. 17 depicts the proposed algorithm for obtaining fuzzy regression models.

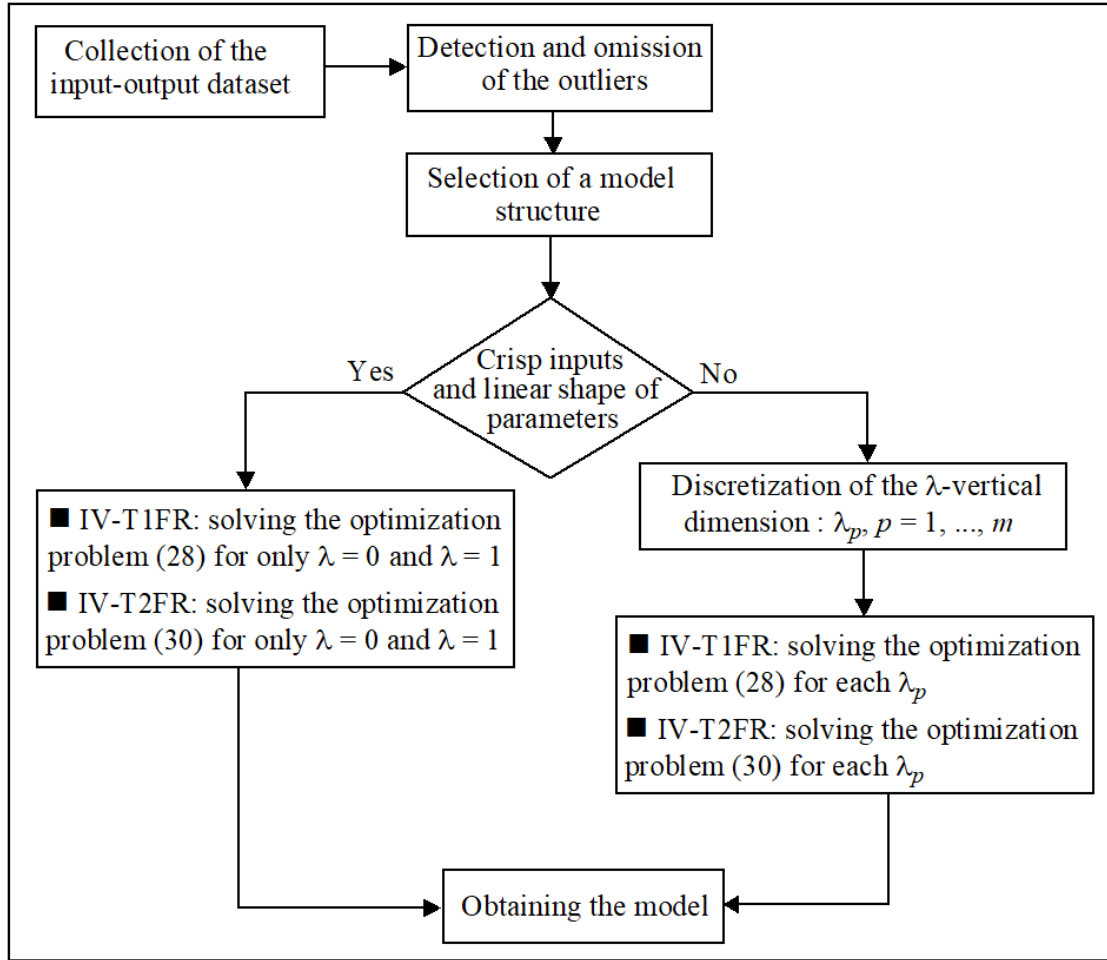


Fig. 17: Proposed interval-valued fuzzy regression methodology

The proposed approach applies to crisp inputs and fuzzy outputs regardless of the shape of IV-T1FSs and IV-T2FSs. However, for interval-valued nonlinear shapes and/or for interval-valued fuzzy inputs, the outputs generated by the regression model become nonlinear interval-valued fuzzy quantities. In this situation, the optimization problem is difficult to solve in a numerical framework. Therefore, the optimization is performed through an interval approach by discretizing the λ -vertical dimension. When considering $\lambda_p, p = 1, \dots, m$ discretized λ -values, the optimization problems (28) and (30) must be solved for each λ_p . In these optimization problems, when IV-fuzzy quantities are required, supplementary monotonicity (consistency) constraints are introduced. For instance, to ensure that $[\theta_j(\lambda)]$ are represented by IV-T1FSs, the following constraint must be imposed:

$$\text{for } p = 1, \dots, m: \forall \lambda_{p+1} \geq \lambda_p : [\theta_j(\lambda_{p+1})] \subseteq [\theta_j(\lambda_p)], j = 0, \dots, M-1$$

In this context, achieving satisfactory results requires the use of a sufficiently small discretization step on λ . This discretization phenomenon leads to a relatively time-consuming optimization procedure. This limitation requires regression practitioners to find a compromise between the computation time and the quality of the approximation of the interval-valued fuzzy sets. To overcome this limitation, most fuzzy regression methodologies assume the common situation where the inputs are crisp and the parameters are triangular or trapezoidal interval-valued fuzzy sets (IV-T1FSs and IV-T2FSs). This assumption implies that the model outputs are also triangular or trapezoidal interval-valued fuzzy sets. In this configuration, the optimization problem can be implemented only at levels $\lambda = 0$ and $\lambda = 1$.

7.2. Interval-valued fuzzy representation

In the proposed application examples, the shape of IV-T1FSs is assumed to be triangular and/or trapezoidal. This particular situation of triangular and/or trapezoidal shapes is widely assumed in fuzzy regression applications that are published in the literature. This assumption is essentially motivated by the ease of computing and interpretation. Therefore, the linear shape (triangular and/or trapezoidal) allows us to express the model output, the parameters, and the optimization problem using only the levels $\lambda = 0$ and $\lambda = 1$. Nevertheless, the proposed method can be applied regardless of the shape of IV-T1FSs (linear piecewise, polynomial, semi-elliptic, ...) under the penalty of heavier computing and a time-consuming optimization procedure [12].

In fuzzy regression approaches, different types of IV-T1FS input-output data have been considered, i.e., crisp-inputs/fuzzy outputs and fuzzy inputs/fuzzy outputs. In the proposed applications, the observed inputs are crisp and the observed outputs are assumed to be triangular IV-T1FSs. In this framework, it has been shown [9] that when assuming trapezoidal IV-T1FS parameters, a fuzzy regression model can always be identified independently of the observed data and the inclusion constraints in the optimization problem can always be satisfied. Hence, the assumptions of crisp inputs and the trapezoidal shape of parameters cause the model output to be trapezoidal.

Usually, the shape of the model output is inherent to the nature and the shape of inputs and parameters. Two different situations can be distinguished according to the nature of inputs. The first situation refers to the case when the inputs are crisp, while the second situation concerns the case where the inputs are fuzzy (IV-T1FSs). Therefore, if the inputs are crisp (the first situation), the model output shape follows the shape of parameters, i.e., if the parameter's shape is linear (respectively nonlinear), the model output is also linear (respectively nonlinear). In the second case where the inputs are fuzzy and represented by IV-T1FSs, the model output shape is always a nonlinear IV-T1FS even though the parameters and inputs have a linear shape (triangular or trapezoidal IV-T1FSs). Indeed, some operations (like the product) between inputs and parameters induce a nonlinearity shape of the model output.

As discussed in section 7.1, the nonlinearity shape of the model output does not allow its formalization using only the levels $\lambda = 0$ and $\lambda = 1$. In the same way, the optimization problem cannot be performed only at levels $\lambda = 0$ and $\lambda = 1$. Therefore, a discretization of λ for all IV-T1FSs is required and the optimization problem is solved for each discretized level λ . In practice and as stated in section 7.1, this design is time-consuming and a compromise between the computation time and the quality of results must be achieved. Although in a nonlinear context, if a linear shape is desired for the model output, it is possible to envelop the model output within a trapezoidal IV-T1FS. However, this approximation operation results in a model and parameters that are too imprecise.

More generally, this paper addresses only IV-T1FSs and IV-T2FSs *via* the GI and TGI concepts, respectively. In this framework, due to the simplicity of their computer coding, CIs are often used as an abstract approximation of crisp sets. A CI is a 1D crisp set, i.e., a special case of a CS. For the same reasons, IV-T1FSs and IV-T2FSs are used as abstract approximations of T1FSs and T2FSs, respectively. Therefore, IV-T1FSs and IV-T2FSs are particular cases of a T1FS and T2FSs, respectively. However, this

regression approach can be extended to address T1FSs and T2FSs through the new concept of thick fuzzy sets [17]. A T1FS, which is denoted \mathbb{X} , can be regarded as a family of nested sets $\mathbb{X}(\lambda)$ according to the vertical dimension λ (λ -cuts), i.e.,

$$\mathbb{X} = \bigcup_{\lambda \in [0,1]} \lambda \mathbb{X}(\lambda)$$

where $\mathbb{X}(\lambda)$ is a set. By adopting the same reasoning, a T2FS can be regarded as a thick fuzzy set that is interpreted as a family of nested thick sets according to the vertical dimension λ (λ -cuts). A thick set is an uncertain set that is denoted by $[\mathbb{X}]$ and defined by two sets \mathbb{X}^{inf} and \mathbb{X}^{sup} , respectively, such as $\mathbb{X}^{\text{inf}} \subset \mathbb{X}^{\text{sup}}$. These sets are regarded as lower and higher bounds of \mathbb{X} . Therefore, the thick set $[\mathbb{X}]$ can be represented as an interval of sets where \mathbb{X}^{inf} and \mathbb{X}^{sup} are its bounds, i.e., $[\mathbb{X}] = [\mathbb{X}^{\text{inf}}, \mathbb{X}^{\text{sup}}]$. The difference $\mathbb{X}^? = \mathbb{X}^{\text{sup}} \setminus \mathbb{X}^{\text{inf}}$ that represents the uncertainty is called the penumbra. Furthermore, an uncertain T1FS is denoted by $[\mathbb{X}]$ (in bold) and defined by two T1FSs \mathbb{X}^{inf} and \mathbb{X}^{sup} , respectively, such as $\mathbb{X}^{\text{inf}} \subset \mathbb{X}^{\text{sup}}$. These T1FSs are regarded as lower and higher bounds of \mathbb{X} . Therefore, the thick fuzzy set $[\mathbb{X}]$ is an interval of T1FSs where \mathbb{X}^{inf} and \mathbb{X}^{sup} are its bounds, i.e.,

$$[\mathbb{X}] = [\mathbb{X}^{\text{inf}}, \mathbb{X}^{\text{sup}}] = \bigcup_{\lambda \in [0,1]} \lambda [\mathbb{X}(\lambda)] = \left[\bigcup_{\lambda \in [0,1]} \lambda \mathbb{X}^{\text{inf}}(\lambda), \bigcup_{\lambda \in [0,1]} \lambda \mathbb{X}^{\text{sup}}(\lambda) \right]$$

Fig. 18 depicts a 2D thick set, a 2D T1FS and a 2D thick fuzzy set.

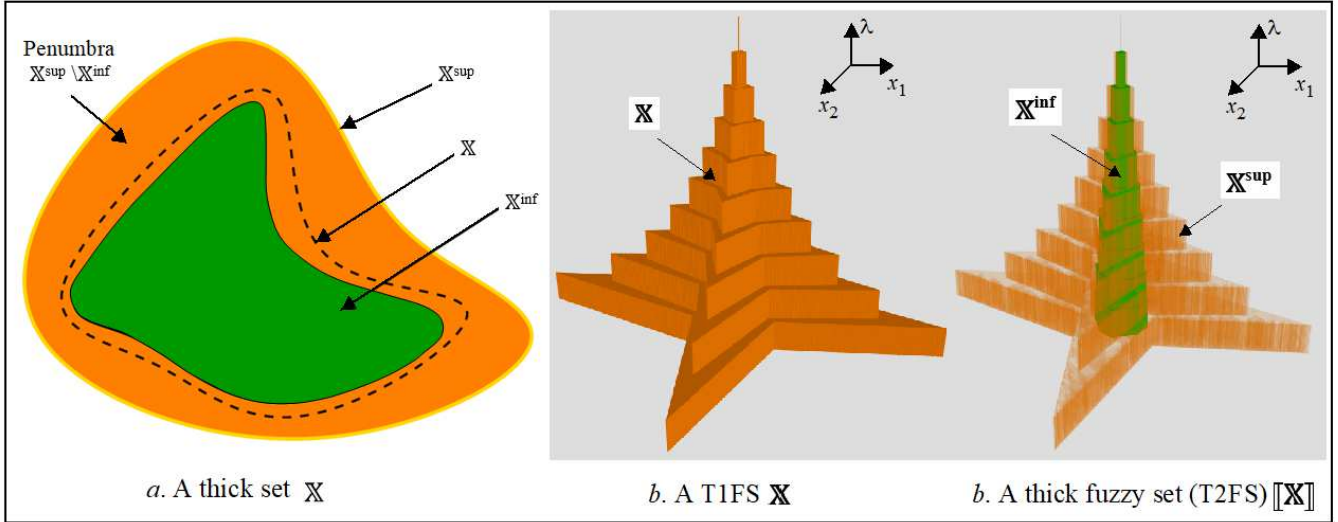


Fig 18: Representation of a 2D thick set, a 2D T1FS and a 2D thick fuzzy set

The uncertainty is exhibited by the penumbra $\mathbb{X}^{\text{sup}} \setminus \mathbb{X}^{\text{inf}}$. According to the λ -cut dimension, the penumbra concept in the thick fuzzy sets representation can be regarded as the footprint of the uncertainty phenomenon in the T2FS view. In the situation when $\mathbb{X}^{\text{inf}} = \mathbb{X}^{\text{sup}} = \mathbb{X}$, the thick fuzzy set $[\mathbb{X}]$ becomes a T1FS \mathbb{X} . In this context, the propagation of T1FSs and T2FSs through regression models is achieved by interval arithmetic over sets and thick sets. Furthermore, this concept of a thick fuzzy set can be applied not only in regression problems but also to more applications in many domains, such as decision-making and linear and nonlinear uncertain fuzzy control. For more details on thick sets and thick fuzzy sets, see [17][29].

7.3. Model robustness, sensitivity and validation

The presence of outliers in the data with fuzzy regression techniques may lead to inaccurate results. Detection to omit or to reduce the impact of such outlier data has an important effect for making rectifications in the obtained model [72]. Therefore, in the presence of detected outliers, two actions are available to us: either they are omitted or filtered [54][55], or they are tolerated while ensuring a reduction of their effect by guaranteeing that the proposed regression methods are robust against these outliers [4].

The choice of either elimination or reduction represents two views to prevent the negative effect of outliers in the adopted fuzzy regression. The outlier problem has been more focused on the detection of outliers rather than their impact on the accuracy of the regression fuzzy model [72]. There are some studies that focus on the use of robust methods for handling this problem [4]. The regression approach addressed in this paper is sensitive to outliers. Therefore, it is necessary to eliminate the outliers before the method implementation to avoid overly imprecise models. For instance, a typical method to treat outliers is to detect and eliminate them using outlier detection algorithms [3][40]. In situations where the outliers are not omitted, it is important to implement methods that are robust against the outliers to make reliable predictions.

Validation and cross-validation methods are essential for evaluating the validity and robustness of the obtained fuzzy regression models. The model validity is judged through indicators such as R^2 and $RMSE$ and others. The data are usually split into a training set for identifying the model and a test set for its validation. A first strategy, which is known as the validation set approach, consists of splitting 50% of the data set for training and the remaining 50% for validation. This approach is relatively severe because when the model is learned on only 50% of the data, it is possible that important phenomena have been forgotten. This situation leads to a higher bias in the validation. To overcome this difficulty, the k -fold cross-validation procedure is often used. In this procedure, the data are randomly split into k -folds. For each k -fold, the $k-1$ folds serve as the training set, and the effectiveness of the model is tested on the k^{th} fold. This operation is repeated until each of the k -folds has served as test data. Generally, a lower value of k is more biased and hence undesirable. On the other hand, a higher value of k is less biased but can suffer from large variability and is time consuming. A compromise for the choice of k must be made by the regression practitioner in relation to the data exploited. Therefore, the smallest value of k leads to a validation set approach. In contrast, the greatest value of k refers to the leave-one-out cross-validation (LOOCV) approach. To judge the validity, sensitivity and robustness of the obtained regression models, two examples have been added in Appendix C, where cross-validation approaches have been employed.

In fuzzy regression, two problems can make the handling and identification of fuzzy regression models more complex. The first problem concerns the evolution of the uncertain trend of the model output. To illustrate this problem, let us consider the case of the first-order linear model (19). According to (19), it can be stated that the radius of the interval output can only increase (respectively, decrease) with the evolution of x if x is positive (respectively, negative):

$$R([\hat{y}(\lambda)]) = R([\theta_0(\lambda)]) + R([\theta_1(\lambda)]) \times |x|; \text{ and } \frac{d[\hat{y}(\lambda)]}{dx} = R([\theta_1(\lambda)]) \times \text{sign}(x)$$

Therefore, the increasing trend of the output radius results in a model that may be excessively uncertain. To overcome this problem, an adjustment parameter for shifting the model inputs to modify the trend of the imprecise output has been proposed in [9]. The second problem refers to the overestimation phenomenon. Therefore, if the proposed unified approach takes advantage of flexibility, rigor and guaranteed results of interval arithmetic and reasoning, it can sometimes be criticized for its accumulation of fuzziness. Because interval arithmetic guarantees the set of all possible results, the pessimistic independence property between the interval-value quantities is implicitly assumed. This phenomenon can cause overestimation of uncertainties in the regression model outputs. This overestimation problem can be reduced by implementing some extensions and hybridizations of interval arithmetic [15].

7.4. Model optimality

In the context of regression, two main problems arise, namely, the model structure specification and the estimation of the given model. Based on a selection procedure, the first problem is focused on the choice of a suitable model structure for a data set. This problem is traditionally addressed a priori to take advantage of the estimators found for the second one. Therefore, the model selection is the process of choosing one among many candidate models. Consequently, it is possible to compare different models to select the most

suitable model. In the literature on information theory, we have some methods capable of quantifying a posteriori the performance of a model and its complexity. Thus, a cost function is often defined to integrate the criterion associated with the model identification and a function exhibiting its complexity. In this context, the AIC [1] and the Rissanen minimum description length criterion (MDL) can be used [71]. In our context of possibilistic fuzzy regression, the AIC has been preferred [12]. The optimal model corresponds to that for which the employed criterion is the minimum. For instance, and as mentioned in section 6, the AIC principle has been employed to select the control intervals (and consequently, the number of piecewise polynomials) to find a balance between the B-spline complexity and its approximation quality. Furthermore, and as discussed in section 4.2, the model choice from the linear/nonlinear or parametric/nonparametric regression model depends on the prior knowledge of the functional form of the relationship \mathcal{M} and the realization of a trade-off between complexity, computational cost and accuracy. In a large body of literature on fuzzy regression analysis, most research has focused on predefined model structures, especially fuzzy linear regression models. In this paper, as commonly used, the model structure is assumed to be known or has been determined a priori. In this case, the fuzzy regression problem is reduced to an estimation problem of the model parameters.

7.5. Model interpretability

The philosophy of the proposed unified approach consists of representing a regression model by analytical expressions in which the inputs, outputs and parameters are represented by IV-T1FSs (GIs) and IV-T2FSs (TGIs). This view has the advantage of being able to integrate and manipulate uncertainty in fuzzy regression models while maintaining the flexibility of interval arithmetic and reasoning. The propagation of GIs and TGIs through the regression models respects the closure property and provides a result of the same nature as the initial information. Therefore, the model output has the same interpretation as inputs and/or parameters in terms of uncertainty interpretations. Fuzzy regression usually exhibits an epistemic interpretation. Therefore, in the IV-T1FR, the inputs, outputs and/or parameters are interpreted as possibility distributions regarded as IV-T1FSs (see section 3.3). In this context, the uncertainty can be quantified thorough possibility and necessity measures. This interpretation can also be accomplished using belief function theory in which a plausibility measure coincides with a possibility measure, and a measure of belief is a measure of necessity (for more details on the interpretation of fuzzy regression using possibility and belief functions theories, see [12]). In the IV-T2FR, the inputs, outputs and parameters are interpreted as uncertain possibility distributions (see section 3.4). Therefore, the exact shape of a possibility distribution is unknown, and only its bounds are available. In this case, several different interpretations can be associated with these bounds according to the domain of application used. For example, they can be interpreted as the higher and lower bounds of uncertainty in some decision problems. They can also be interpreted as representations of extreme situations in adverse and favorable situations.

8. Conclusions

In this paper, a unified approach to address type-1 and type-2 fuzzy regressions has been described (IV-T1FR and IV-T2FR). The proposed work extends the CIR to a fuzzy context to improve its flexibility and specificity, revises some methodological and philosophical reflections on the interest, representativeness and typology of interval-valued fuzzy regression and makes it possible to address T1FR and the IV-T2FR in a single formalism. Therefore, the IV-T2FR is an uncertain IV-T1FR, i.e., in the absence of uncertainty about the IV-T1FS shape, the IV-T2FR becomes an IV-T1FR. Our approach is based on the use of GIs and TGIs for representing inputs, outputs and/or parameters of regression models. Furthermore, GIs and TGIs can be regarded as IV-T1FSs and IV-T2FSs, respectively.

The proposed methodology is not a statistical approach and takes advantage of the flexibility, rigor and guaranteed results of interval arithmetic and reasoning for regression model construction and for the propagation of information through these models. The IV-T1FR and IV-T2FR methodologies are illustrated using application examples. In addition to regression issues, several other potential applications of our

approach, and more particularly, the TGI concept, can be envisioned in modeling, decision-making problems, etc. In this paper, the suggested approach addresses only IV-T1FSs and IV-T2FSs. However, it can be extended to T1FSs and T2FSs according to the concepts of thick fuzzy sets (TFS) [17]. In the short term, work will be devoted to these interesting extensions and their implantations through real-world applications. In the long term, we will focus on the development of methodologies to explain and to interpret the behavior and performances of deep learning systems using fuzzy models. It is well known that the driving forces behind deep learning success are due to large-scale training data and better parameter learning strategies. Therefore, deep learning approaches have poor transparency causing difficulty in their understanding and interpretation. Moreover, the effectiveness of these deep techniques can be altered when the data is not well defined due to the presence of uncertainty, ambiguity, vagueness, and incompleteness. The goal is to go towards deep fuzzy regression models where the objective is to bring together the strong learning capacity of deep learning techniques (based on neural networks) with the high level of interpretability and transparency of fuzzy models. For instance, to model a system, it can be envisioned that basic blocks (capsules) can be locally learned using deep techniques. The interconnection architecture of these capsules is then carried out through a fuzzy model where uncertainty, vagueness, ambiguity, and incompleteness of the data and the overall system functioning can be taken into account. To achieve mature results, this perspective will require a decade of research and development.

References

- [1] Akaike H., A look at the statistical model identification, *IEEE Trans. on Auto. Control*, Vol. 19, N° 6, pp. 716-723, 1974.
- [2] Bajestani N-S., Kamyad A-V. and Zare A., A Piecewise Type-2 Fuzzy Regression Model, *Int. J. of Comput. Intelligence Systems*, Vol. 10, pp. 734–744, 2017.
- [3] Chachi J., Roozbeh M., A fuzzy robust regression approach applied to bedload transport data, *Comm. Statist. Simulation Comput.* Vol. 46, pp. 1703–1714, 2017.
- [4] Chakravarty S., Demirhan H., Baser F., Fuzzy regression functions with a noise cluster and the impact of outliers on mainstream machine learning methods in the regression setting, *Applied Soft Computing*, Vol. 96, pp. 1-17, 2020.
- [5] Coppi R., D’Urso P., Giordani P., Santoro A., Least squares estimation of a linear regression model with LR fuzzy response, *Computational Stat. & Data Analysis*, Vol. 51 (1), pp. 267-286, 2006.
- [6] Coppi R., Management of uncertainty in statistical reasoning: The case of regression analysis, *Int. J. Approx. Reason.*, Vol. 47(3), pp. 284–305, 2008.
- [7] Blanco-Fernandez A., Colubi A., Garcia-B M., A set arithmetic-based linear regression model for modeling interval-valued responses through real-valued variables, *Inform. Sci.*, Vol. 247, pp. 109-122, 2013.
- [8] Bissierier A., Boukezzoula R., Galichet S., Linear Fuzzy Regression Using Trapezoidal Fuzzy Intervals, *Journal of Uncertain Systems*, Vol. 4, N° 1, pp. 59-72, 2010.
- [9] Bissierier A., Boukezzoula R., Galichet S., A Revisited Approach for Linear Fuzzy Regression Using Trapezoidal Fuzzy Intervals, *Inform. Sci.*, Vol. 180, N°. 19, pp. 3653-3673, 2010.
- [10] Boukezzoula R., Galichet S., Bissierier A., A Midpoint–Radius approach to regression with interval data, *International Journal of Approx. Reason.*, Vol. 52, N° 9, pp. 1257-1271, 2011.
- [11] Boukezzoula R., Galichet S., Foulloy L., Elmasry M., Extended gradual interval (EGI) arithmetic and its application to gradual weighted averages, *Fuzzy Sets & Syst.*, Vol. 257, pp. 67-84, 2014.
- [12] Boukezzoula R., Galichet S. and Coquin D., From fuzzy regression to gradual regression: Interval-based analysis and extensions, *Inform. Sci.*, Vol. 441, pp. 18-40, 2018.
- [13] Boukezzoula R., Galichet S., Foulloy L., Min and Max Operators for Gradual intervals, *IEEE Trans. on Fuzzy Systems*, Vol. 26 (6), pp. 3569-3578, 2018.
- [14] Boukezzoula R., Jaulin L. and Foulloy L., Thick gradual intervals: An alternative interpretation of type-2 fuzzy intervals and its potential use in type-2 fuzzy computations, *Eng. Appl. of Artificial Intelligence*, Vol. 85, pp. 691-712, 2019.
- [15] Boukezzoula R., Foulloy L., Coquin D. *et al.* Gradual interval arithmetic and fuzzy interval arithmetic. *Granul. Comput.*, <https://doi.org/10.1007/s41066-019-00208-z>, 2019.
- [16] Boukezzoula R., Coquin D., A decision-making computational methodology for a class of type-2 fuzzy intervals: An interval-based approach, *Information Sciences*, Vol. 510, pp. 256-282, 2020.
- [17] Boukezzoula R., Jaulin L., Desrochers B., Coquin D., Thick Fuzzy Sets (TFSs) and Their Potential Use in Uncertain Fuzzy Computations and Modeling, accepted in *IEEE Transactions on Fuzzy systems*, DOI 10.1109/TFUZZ.2020.3018550, 2020.
- [18] Cerny M. and Hladik M., Possibilistic linear regression with fuzzy data: Tolerance approach with prior information, *Fuzzy Sets & Syst.*, Vol. 304, pp. 127-144, 2018.

- [19] Chakraborty C., Chakraborty D., A theoretical development on a fuzzy distance measure for fuzzy numbers, *Mathematical and Computer Modelling*, Vol. 43, pp. 254–261, 2006.
- [20] Chen L-H., Nien S-H., A new approach to formulate fuzzy regression models, *Appl. Soft Comp. J.*, Vol. 86, pp. 1-10, 2020.
- [21] Chen Y.-S., Fuzzy Ranking and quadratic Fuzzy regression, *Comput. Math. Appl.*, Vol. 38 (11–12), pp. 265–279, 1999.
- [22] Chernozhukov V., Hong H., and Tamer E., Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, Vol. 75 (5), pp. 1243–1284, 2007.
- [23] Chuang C.C., Extended support vector interval regression networks for interval input–output data, *Inform. Sci.*, Vol. 178 (3), pp. 871-891, 2008.
- [24] Chukhrova N., Johannssen A., Fuzzy regression analysis: Systematic review and bibliography, *Applied Soft Computing*, Vol. 84, pp. 1-29, 2019.
- [25] Cuso I. and Dubois D., Statistical reasoning with set-valued information: Ontic vs. epistemic views, *Int. J. of Approx. Reas.*, Col. 55, N° 7, pp. 1502-1518, 2014.
- [26] De Boor C., On calculating with B-splines, *J. Approx. Theory*, Vol. 6, pp. 50–62, 1972.
- [27] De Boor C., *A Practical Guide to Splines*, Springer-Verlag, N-Y, 1978.
- [28] De Lima Neto E.-A., De Carvalho F.A.T., Nonlinear regression applied to interval-valued data, *Pattern Anal. Appl.*, Vol. 20, pp. 809–824, 2017.
- [29] Desrochers B., and Jaulin L., Thick set inversion, *Artificial Intelligence*, Vol. 249, pp. 1-18, 2017.
- [30] Diamond P., Fuzzy Least Squares, *Inform. Sci.* 46. pp. 141-157, 1988.
- [31] Dubois D., Kerre E., Mesiar R. and Prade H., Fuzzy interval analysis, in: D. Dubois, H. Prade (Eds.), *Fundamentals of Fuzzy Sets*, The Handbooks of Fuzzy Sets Series, Kluwer, Boston, pp. 483-581, 2000.
- [32] Dubois D., On various ways of tackling incomplete information in statistics, *Int. J. of Approx. Reason.*, Vol. 55 (7), pp. 1570-1574, 2014.
- [33] D’Urso P. and Gastaldi T., A least-squares approach to fuzzy linear regression analysis, *Comp. Stat. & Data Anal.*, 34, pp. 427-440, 2000.
- [34] D’Urso P., Giordani P., A least squares approach to principal component analysis for interval valued data, *Chemometrics and Intelligent Laboratory Systems*, Vol. 70, pp. 179-192, 2004.
- [35] D’Urso P. and M. Leski J-M., Fuzzy c-ordered medoids clustering for interval-valued data, *Pattern Recognition*, Vol. 58, pp. 49-67, 2016.
- [36] Dymova L., Sevastjanov P., and Tikhonenko A., An interval type-2 fuzzy extension of the TOPSIS method using alpha-cuts, *Knowledge-Based Systems*, Vol. 83, pp. 116–127, 2015.
- [37] Figueroa-Garcia J.C, Chalco-Cano Y., Roman-Flores H., Distance measures for Interval Type-2 fuzzy numbers, *Discrete Applied Mathematics*, Vol. 197 pp. 93–102, 2015.
- [38] Ferson S. et al., *Experimental uncertainty estimation and statistics for data having interval uncertainty*, Sandia Nat. Lab., SAND 2007.
- [39] Fortin J., Dubois D. and Fargier H., Gradual numbers and their application to fuzzy interval analysis, *IEEE Trans. on Fuzzy Syst.*, Vol. 16 (2), pp. 388-402, 2008.
- [40] Ghafoori Z., Erfani S.M., Rajasegarar S., Bezdek J.C., Karunasekera S., Leckie C., Efficient unsupervised parameter estimation for one-class support vector machines, *IEEE Trans. Neural Netw. Learn. Syst.* Vol. 29, pp. 5057–5070, 2018.
- [41] Gil M.A., Gonzalez-Rodriguez G., Colubi A. and Montenegro M., Testing linear independence in linear models with interval-valued data, *Comput. Stat. and Data Anal.*, Vol. 51 (6), pp. 3002-3015, 2007.
- [42] Gonzalez-Gonzalez D.S., Praga Alejo R.J., Cantu-Sifuentes M., Luis M. Torres-Trevino L.M., Méndez G.M., A non-linear fuzzy regression for estimating reliability in a degradation process, *Applied Soft Computing*, Vol. 16, pp. 137-147, 2014.
- [43] Guo P. and Tanaka H., Dual models for possibilistic regression analysis, *Comp. Stat. & Data Anal.* Vol. 51 (1), pp. 253-266, 2006.
- [44] Hao P.Y., Interval regression analysis using support vector networks, *Fuzzy Sets & Sys.*, Vol. 160, pp. 2466-2485, 2009.
- [45] Hao P.Y., Pairing support vector algorithm for data regression, *Neurocomputing*, Vol. 225, pp. 174–187, 2017.
- [46] Hao P.Y., Dual possibilistic regression analysis using support vector networks, *Fuzzy Sets and Systems*, doi.org/10.1016/j.fss.2019.03.012, 2019.
- [47] Harmening C., and Neuner H., Choosing the Optimal Number of B-spline Control Points (Part 1: Methodology and Approximation of Curves), *Journal of Appl. Geodesy*, Vol. 10 (3), pp.139-157, 2016.
- [48] Hayashi I., Tanaka H., The Fuzzy GMDH algorithm by possibility models and its application, *Fuzzy Sets and Systems*, Vol. 36 (2), pp. 245–258, 1990.
- [49] Hesamian G., Akbari M.G., A fuzzy additive regression model with exact predictors and fuzzy responses, *Applied Soft Computing*, Vol. 95, 106507, 2020.
- [50] Hladik M. and Cerny M., Interval regression by tolerance analysis approach, *Fuzzy Sets & Sys.* Vol. 193, pp. 85-107, 2012.
- [51] Hong D.H., Song J.-K. and Young H., Fuzzy least-squares linear regression analysis using shape preserving operations, *Inform. Sci.*, Vol. 138, pp. 185-193, 2001.

- [52] Hong D.H. and Hwang C., Extended fuzzy regression models using regularization method, *Inform. Sci.*, Vol. 164, pp. 31-46, 2004.
- [53] Hosseinzadeh E., Hassanpour H., Arefi M., A weighted goal programming approach to fuzzy linear regression with crisp inputs and type-2 fuzzy outputs, *Soft Comput.* Vol. 19 (5), pp. 1143-1151, 2014.
- [54] Hu J., Zheng K., A novel support vector regression for data set with outliers, *Applied soft computing*, Vol. 31, pp. 405-411, 2015.
- [55] Hung W-L., Yang M-S., An omission approach for detecting outliers in fuzzy regression models, *Fuzzy sets and Systems*, Vol. 157, pp. 3109-3122, 2006.
- [56] Ismail O., Interval B-Spline Curve Fitting, *Int. Journal of Video & Image Process. and Network Security*, Vol. 14 (6), pp. 1-7, 2014.
- [57] Jaulin L. et al., *Applied Interval Analysis, With Examples in Parameter and State Estimation, Robust Control and Robotics*. New York: Springer-Verlag, 2001.
- [58] Jhong J-H., Koo J-Y., Lee S-W., Penalized B-spline estimator for regression functions using total variation penalty, *Journal of Stat. Plan. and Infer.* Vol. 184, pp. 77-93, 2017.
- [59] Kulpa Z., Diagrammatic representation for interval arithmetic, *Linear Alg. and its Appl.*, Vol. 324, N°1-3, pp. 55-80, 2001.
- [60] Lee H., Tanaka H., Fuzzy approximations with non-symmetric Fuzzy parameters in Fuzzy regression analysis, *Jour. Oper. Res. Soc.*, Vol. 42 (1), pp. 98-112, 1999.
- [61] Lodwick W.A. and Dubois D., Interval linear systems as a necessary step in fuzzy linear systems, *Fuzzy Sets & Syst.*, Vol. 281, pp. 227-251, 2015.
- [62] Luo R., Qi X., Sparse wavelet regression with multiple predictive curves, *Journal of Multi. Anal.*, Vol. 134, pp. 33-49, 2015.
- [63] Mendel J. M. and R. I. John, Type-2 fuzzy sets made simple, *IEEE Trans. on Fuzzy Syst.*, Vol. 10, N° 2, pp. 117-127, 2002.
- [64] Mendel J.M., John R.I. and Liu F., Interval type-2 fuzzy logic systems made simple, *IEEE Trans. Fuzzy Syst.* Vol. 14 (6), pp. 808-821, 2006.
- [65] Mendel J.M., *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper Saddle River, NJ: Prentice-Hall, 2001.
- [66] Moore R. E., *Interval Analysis*, P-Hall, Englewood Cliffs, N.J., 1966.
- [67] Peng X., Chen D., PTSVRs: Regression models via projection twin support vector machine, *Inform. Sci.*, Vol. 435, pp. 1-14, 2018.
- [68] Pekala B., Comparing uncertainty data in epistemic and ontic sense used to decision making problem, *Iranian Journal of Fuzzy Systems*, Vol. 16 (2), pp. 45-55, 2019.
- [69] Poleshchuk O., E.Komarov E., A Fuzzy Nonlinear Regression Model for Interval Type-2 Fuzzy Sets, *Int. J. of Mathematical and Comput. Sciences*, Vol. 8 (6), pp. 840-844, 2014.
- [70] Kocadagli O., A new approach for Fuzzy multiple regression with Fuzzy output, *Int. J. Ind. Syst. Eng.*, Vol. 9 (1) pp. 1748-5037, 2011.
- [71] Rissanen J., Modeling by shortest data description. *Automatica*, Vol. 14, N° 5, pp. 465-471, 1978.
- [72] Shakouri H.G., Nadimi R., Outlier detection in fuzzy linear regression with crisp input-output by linguistic variable view, *Applied soft computing*, Vol. 13, pp. 734-742, 2013.
- [73] Tanaka H., Uejima S., Asai K., Linear regression analysis with fuzzy model, *IEEE Trans. SMC-2*, pp. 903-907, 1982.
- [74] Tanaka H., Hayashi I., Watada J., Possibilistic linear regression analysis for fuzzy data, *EJOR*, 40, pp. 389-396, 1989.
- [75] Tanaka H. and Ishibushi H., Identification of possibilistic linear systems by quadratic membership functions of fuzzy parameters, *Fuzzy Sets & Syst.* 41, pp. 145-160, 1991.
- [76] Trindade R. M. P., Bedregal B. R. C., Neto A. D. D., Acioly B. M., *An Interval Metric*, New Advanced Technologies, Edited by Aleksandar Lazinica, pp. 323-340, 2010.
- [77] Tsybakov A.B., *Introduction to Nonparametric Estimation*. In: Springer Series in Statistics, 2009.
- [78] Tuohy S.T., Maekawa T., Shen G., Patrikalakis N.M., Approximation of measured data with interval B-splines, *Computer-Aided design*, Vol. 29 (11), pp. 791-799, 1997.
- [79] Wei Y., Watada J., Building a Type II Fuzzy Qualitative Regression, *Intel. Dec. Technol.* Vol. 1 (15), pp. 145-154, 2012.
- [80] Xu Q., et al., Composite quantile regression neural network with applications, *Expert Syst. with Appl.*, Vol. 76, pp. 129-139, 2017.
- [81] Yu J.-R., Lee C.-W., Piecewise regression for Fuzzy input-output data with automatic change-point detection by quadratic programming, *Appl. Soft Comput.*, Vol. 10 (1), pp. 111-118, 2010.
- [82] Zadeh L.A., The concept of a linguistic variable and its application to approximate reasoning, *Inf. Sci.*, Vol. 8, pp. 199-250, pp. 301-357, pp. 9:42-80, 1975.
- [83] Zhao G., Ma Y., Robust nonparametric kernel regression estimator, *Stat. and Probability Letters*, Vol. 116, pp. 72-79, 2016.

Appendix A

The set $I = \{[a^-, a^+] \mid a^- \leq a^+, a^-, a^+ \in \mathfrak{R}\}$ denotes the set of CIs, and \mathbb{O} represents the subset of CIs in I that contain zero in their interiors. When considering two CIs $[a] = [a^-, a^+]$ and $[b] = [b^-, b^+]$, the four SIA operations are defined by the following expressions [66]:

$$\text{Addition: } \forall [a], [b] \in \mathbb{I}: [a] + [b] = [a^- + b^-, a^+ + b^+]$$

$$\text{Subtraction: } \forall [a], [b] \in \mathbb{I}: [a] - [b] = [a^- - b^+, a^+ - b^-]$$

$$\text{Multiplication: } \forall [a], [b] \in \mathbb{I}: [a] \times [b] = [\min(\Phi), \max(\Phi)]; \quad \Phi = \{a^-b^-, a^-b^+, a^+b^-, a^+b^+\}$$

$$\text{Division: } \forall [a] \in \mathbb{I}, [b] \in \mathbb{I} \setminus \mathbb{O}: [a] \div [b] = [a] \times (1 \div [b]); \text{ where } 1 \div [b] = [1/b^+, 1/b^-]$$

If $0 \in [b]$, then $[a] \div [b] = \mathfrak{R}$. The SIA operations on CIs can be directly extended to TIs [16][29], i.e.,

$$\text{Addition: } \forall [[a]], [[b]] \in \mathbb{I}: [[a]] + [[b]] = [[a^-, [a^+]] + [[b^-, [b^+]]] = [[a^- + b^-, [a^+ + b^+]]]$$

$$\text{Subtraction: } \forall [[a]], [[b]] \in \mathbb{I}: [[a]] - [[b]] = [[a^-, [a^+]] - [[b^-, [b^+]]] = [[a^- - b^+, [a^+ - b^-]]]$$

$$\text{Multiplication: } \forall [[a]], [[b]] \in \mathbb{I}: [[a]] \times [[b]] = [[a^-, [a^+]] \times [[b^-, [b^+]]] = [[\min(\Phi), \max(\Phi)]]$$

where $\Phi = \{[a^-] \times [b^-], [a^-] \times [b^+], [a^+] \times [b^-], [a^+] \times [b^+]\}$. The *min* and *max* operators between two CIs $[a] = [a^-, a^+]$ and $[b] = [b^-, b^+]$ are expressed as follows:

$$\min([a], [b]) = [\min(a^-, b^-), \min(a^+, b^+)]; \quad \max([a], [b]) = [\max(a^-, b^-), \max(a^+, b^+)]$$

$$\text{Division: } \forall [[a]] \in \mathbb{I}, [[b]] \in \mathbb{I} \setminus \mathbb{Z}: [[a]] \div [[b]] = [[a^-, [a^+]] \div [[b^-, [b^+]]] = [[a]] \times (1 \div [[b]])$$

$$\text{where } 1 \div [[b]] = 1 \div [[b^-, [b^+]]] = [[1 \div b^+, 1 \div b^-]]$$

Each operation $\odot \in \{+, -, \times, \div\}$ between TIs is interpreted as follows: if $[a] \in [[a]]$ ($a^- \in [a^-]$ and $a^+ \in [a^+]$) and $[b] \in [[b]]$, then the interval $[x] = [a] \odot [b]$ satisfies $[x] \in [[a]] \odot [[b]]$. Furthermore, each operation $[a] \odot [b]$ is performed via SIA operations.

Appendix B

The set Γ denotes the set of GI and $\Gamma \mathbb{O}$ the subset of GIs in Γ that contain zero in their interiors. When considering two GIs $[a(\lambda)] = [a^-(\lambda), a^+(\lambda)]$ and $[b(\lambda)] = [b^-(\lambda), b^+(\lambda)]$, the SIA operations on CIs can be extended to GIs (see [11][12][16]), where the CIs are replaced by GIs, i.e.,

$$\text{Addition: } \forall [a(\lambda)], [b(\lambda)] \in \mathbf{GI}: [a(\lambda)] + [b(\lambda)] = [a^-(\lambda) + b^-(\lambda), a^+(\lambda) + b^+(\lambda)]$$

$$\text{Subtraction: } \forall [a(\lambda)], [b(\lambda)] \in \mathbf{GI}: [a(\lambda)] - [b(\lambda)] = [a^-(\lambda) - b^+(\lambda), a^+(\lambda) - b^-(\lambda)]$$

$$\text{Multiplication: } \forall [a(\lambda)], [b(\lambda)] \in \mathbf{GI}: [a(\lambda)] \times [b(\lambda)] = [\min(\Phi(\lambda)), \max(\Phi(\lambda))]$$

$$\text{where: } \Phi(\lambda) = \{a^-(\lambda)b^-(\lambda), a^-(\lambda)b^+(\lambda), a^+(\lambda)b^-(\lambda), a^+(\lambda)b^+(\lambda)\}$$

$$\text{Division: } \forall [a(\lambda)] \in \mathbf{GI}, [b(\lambda)] \in \mathbf{GI} \setminus \mathbb{Z}: [a(\lambda)] \div [b(\lambda)] = [a(\lambda)] \times (1 \div [b(\lambda)]); \text{ with } 1 \div [b(\lambda)] = [1/b^+(\lambda), 1/b^-(\lambda)]$$

As an IV-TIFS is a special case of a GI, the operations on IV-TIFSs are performed using operations over GIs. As the SIA over CIs is extended to TIs, the SIA using GIs can be extended to TGIs [16], i.e.,

$$\text{Addition: } \forall [[a(\lambda)]], [[b(\lambda)]] \in \mathbf{GI}: [[a(\lambda)]] + [[b(\lambda)]] = [[a^-(\lambda)] + [b^-(\lambda)], [a^+(\lambda)] + [b^+(\lambda)]]$$

$$\text{Subtraction: } \forall [[a(\lambda)]], [[b(\lambda)]] \in \mathbf{GI}: [[a(\lambda)]] - [[b(\lambda)]] = [[a^-(\lambda)] - [b^+(\lambda)], [a^+(\lambda)] - [b^-(\lambda)]]$$

$$\text{Multiplication: } \forall [[a(\lambda)]], [[b(\lambda)]] \in \mathbf{GI}: [[a(\lambda)]] \times [[b(\lambda)]] = [[\min(\Phi(\lambda)), \max(\Phi(\lambda))]]$$

$$\text{where: } \Phi(\lambda) = \{[c_1(\lambda)], [c_2(\lambda)], [c_3(\lambda)], [c_4(\lambda)]\}; \text{ with:}$$

$$[c_1(\lambda)] = [a^-(\lambda)] \times [b^-(\lambda)], [c_2(\lambda)] = [a^-(\lambda)] \times [b^+(\lambda)], [c_3(\lambda)] = [a^+(\lambda)] \times [b^-(\lambda)]; [c_4(\lambda)] = [a^+(\lambda)] \times [b^+(\lambda)]$$

As for CIs, the *min* and *max* between two GIs $[a(\lambda)]$ and $[b(\lambda)]$ are expressed as follows:

$$\min([a(\lambda)], [b(\lambda)]) = [\min(a^-(\lambda), b^-(\lambda)), \min(a^+(\lambda), b^+(\lambda))];$$

$$\max([a(\lambda)], [b(\lambda)]) = [\max(a^-(\lambda), b^-(\lambda)), \max(a^+(\lambda), b^+(\lambda))];$$

$$\text{Division: } \forall [[a(\lambda)]] \in \mathbf{GI}, [[b(\lambda)]] \in \mathbf{GI} \setminus \mathbb{Z}: [[a(\lambda)]] \div [[b(\lambda)]] = [[a(\lambda)]] \times (1 \div [[b(\lambda)]])$$

$$\text{where } 1 \div [[b(\lambda)]] = 1 \div [[b^-(\lambda)], [b^+(\lambda)]] = [[1 \div b^+(\lambda)], 1 \div b^-(\lambda)]]$$

It can be stated that in this extension, all the TIs are replaced by TGIs. However, the multiplication and division operations, where *min* and *max* operations between GIs are needed, deserve more attention. Therefore, compared with CIs where only a single horizontal dimension is considered, GIs add an additional vertical dimension. In this framework, attention must be paid to the points of intersection between ascending (descending) profiles (see [13] for more details for computing *min* and *max* operations between GIs).

Appendix C: Examples of comparison and cross-validation

Two benchmark examples are presented in this appendix. Due to the complexity of the data, an NP interval B-spline model is considered. For comparison purposes with crisp statistical methods, in our implemented method, the midpoint of the interval model output is taken as the best estimation.

1. Example 1

The proposed method is evaluated using the motorcycle benchmark data set, which is commonly used in the testing of regression and machine learning algorithms. This motorcycle data set is heteroscedastic and forms a challenging test case. It is a real-life example where the spread of the noise is correlated with the input values. The data sets used for comparison and the evaluation criteria are similar to those used in [67]. Table 3 lists the comparative learning results between the proposed method and the support vector regression (SVR) technique and its extensions. For more details on these methods, see [45][67]. As in [67], the results are obtained using the leave-one-out (loo) cross-validation method because of the small size of the data set. Specifically, at each run, we use one sample as the test sample and the others as training samples. The performance is evaluated through the SSE/SST and SSR/SST criteria [67].

Table 3 regroups the results extracted from [67] with the results obtained with the B-spline possibilistic approach proposed. According to Table 3, our method yields comparable SSE/SST and SSR/SST values to the other regression methods, which validates the B-spline possibilistic methodology. Notably, a small SSE/SST means good agreement between the estimations and data. However, a smaller SSE/SST is usually counterbalanced by an increased SSR/SST. As detailed in [45][67], extremely small values of SSE/SST are in fact not good, since they likely mean the overfitting of the regressor. A good regressor should provide a compromise between SSE/SST and SSR/SST.

	SSE/SST ^{loo}	SSR/SST ^{loo}
Our Method	0.2283	0.7897
Single-shifted Projection Twin SVR (SPTSVR)	0.2297	0.7729
Pair-shifted Projection Twin SVR (PPTSVR)	0.2287	0.7817
Twin SVR (TWSVR)	0.2305	0.7487
Twin Projection SVR (TPSVR)	0.2331	0.7692
Twin SVR (TSVR)	0.2313	0.7627
SVR	0.2310	0.7586

Table 3: Comparative results on the motorcycle benchmark dataset

2. Example 2

The data are generated by the following expression (with a heteroscedastic characteristic) [80]:

$$y = f(x) + \sigma(x) \cdot \varepsilon; f(x) = (1 - x + 2x^2) \exp(-0.5x^2)$$

where $x \sim U(-4, 4)$, ε is a random error, and $\sigma(x) = (1 + 0.2x)/5$. The same protocol proposed in [80] is adopted here for comparison. Furthermore, to evaluate the robustness of our approach, a random error $N(0, 0.25)$ is used. For each sample, 400 observations are generated. The data are randomly divided into two parts of equal size (200 observations per part): estimation and prediction (validation) parts. The results are evaluated on 999 repetitions to show the performance of the used models. Based on the AIC, the number of

control intervals is chosen to be equal to 15. Our approach has been compared with multiple regression approaches that are reported in [80] (see Table 4). Both estimation (in-sample) and prediction (out-of-sample) performances are evaluated through the mean absolute error (MAE) and the RMSE, as proposed in [80] and reproduced in Table 4. The obtained results indicate that even though these methods are different in their designs, the performance of our approach is efficient compared to most of the techniques in Table 4. The performances of our method are illustrated via the boxplots of MAE and RMSE indicators for both in-sample and out-of-sample (see Fig. 19).

	<i>In sample</i>		<i>Out of sample</i>	
	MAE	RMSE	MAE	RMSE
Our method: $n = 15$	0.042	0.056	0.043	0.060
CQRNN (Composite Quantile Regression Neural Network): $k = 19$	0.040	0.056	0.042	0.057
QRNN (Quantile Regression Neural Network)	0.090	0.125	0.094	0.130
KQR (Kernel Quantile Regression)	0.043	0.059	0.283	0.358
ANN (Artificial Neural Network)	0.178	0.243	0.809	0.991
A-CQR (Additive CQR)	0.238	0.292	0.240	0.295
SI-CQR (Single-Index CQR)	0.305	0.379	0.308	0.382

Table 4: Comparative results (the mean of 999 repetitions)

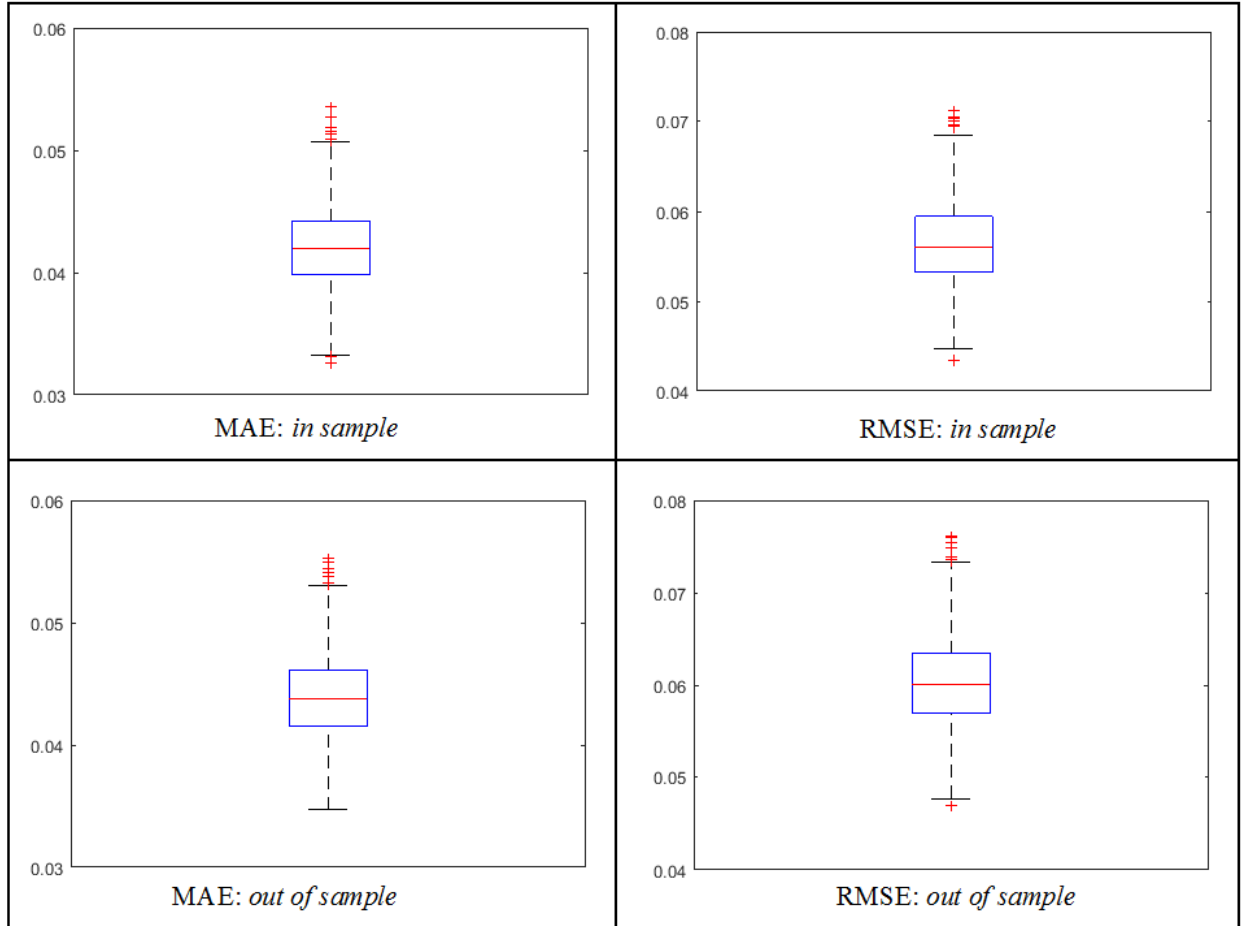


Fig. 19: Boxplots of the indicators MAE and RMSE using the proposed method

This example shows that, in its philosophy, the interval possibilistic regression can be seen as a useful and/or comparable solution in relation to a statistical view. Therefore, we have illustrated that by simply taking the midpoint of the B-spline possibilistic model, the results obtained are comparable to those obtained in [80].