Toward the use of quantile fuzzy transforms for the construction of fuzzy association rules

Nicolás Madrid
Dept. of Applied Mathematics
University of Málaga
Málaga, Spain
nicolas.madrid@uma.es

***Abstract*—This paper analyzes the possibility of defining fuzzy association rules by means of direct quantiles F-transforms. The set of fuzzy association rules is used in a fuzzy inference system, defined by means of the inverse quantile F-transform. The obtained inference system reminds the Takagi-Sugeno one due to the use of a weighted sum to perform the inference. However, there is an important difference: the output is a fuzzy set and, as a result, we require the use of a defuzzification procedure. In addition, in this paper we prove experimentally that the fuzzy set obtained as the output of the proposed inference system is related to a probability distribution.**

***Index Terms*—Fuzzy Transforms, Quantile regression, Fuzzy association rules, Fuzzy inference systems.**

I. Introduction

Fuzzy inference systems have had a significant influence in Engineering during the last 40 years, for instance, in the development of control systems [1]. Currently, it still has the interest of the research community for further developments, both applied and theoretical [2], [3]. The approaches of Mamdani in [8] and Takagi-Sugeno in [9] are the most applied fuzzy inference systems for Engineering tasks, despite of the existence of several other areas concerning inferences in the fuzzy setting, as Fuzzy Logic [4], [6], Fuzzy relation equations [5] or Fuzzy answer set programming [7]. In this paper we focus on the construction of fuzzy association rules for a Takagi-Sugeno type inference system (TS-inference systems).

In the literature, the reader can find a huge variety of methods for the automatic construction of fuzzy rules for TS-inference systems. Such a methods can be classified in different groups according to the used techniques [10]; e.g., genetic based methods [11], neural-network methods [12], [13] or clustering methods [14], [15], among others. The method proposed in this paper is framed in the group of statistics and optimization based approaches [16], [17].

In this paper we propose the construction of a fuzzy inference system similar to the one proposed by Takagi and Sugeno, in the sense that the inference is computed by a weighted sum, but with an important difference: the result of the inference is a fuzzy set. That implies the necessity of considering a defuzzification procedure at the end of the inference. In the literature we can find different extensions of Sugeno inference systems, as those based on intervals [2], type-2 fuzzy sets [20] or intuicionistic fuzzy sets [21]. However, the inference system proposed in this paper does not consider generalization of fuzzy sets, its novelty resides on the use of quantiles F-transforms [22]–[24] for both, the definition of association rules and the inference engine.

This is not the first approach dealing with the use of F-transforms for the construction of fuzzy rules in fuzzy inference systems [25]–[27]. Actually, there are evident relationships between the two kinds of F-transforms, the direct and the inverse, and T-S inference systems. Specifically, the interpolation obtained by the inverse (standard) F-transforms can be considered an inference procedure where the knowledge database is constructed by the direct (standard) F-transforms [25], [27]. Such a construction coincides with the proposal given in [16], in a context of T-S inference systems. This paper goes a step further by proposing the use of quantiles F-transforms [24] for this task. In this way we propose both, a fuzzy inference engine based on the inverse quantile F-transforms and a construction method to create the fuzzy rules based on direct quantile F-transforms. As a remarkable feature, the output obtained by the proposed inference system can be related to quantile regression [18] and to probability distributions conditioned to the satisfiability of fuzzy events [19].

For the sake of a better understating of the proposed fuzzy inference system, in Section II we recall the notions of (standard) F-transform, quantiles F-transforms and enumerate some of their most important properties. Then, in Section III we present the proposed fuzzy inference system and the proposed construction of rules. In Section IV, we show some experiments in both synthetic and real data to illustrate the convenient properties of the approach. Finally in Section V we provide some conclusions and future lines of research.

II. Preliminaries

A. Fuzzy partitions

Let us begin by recalling that a fuzzy set A on a universe $\mathscr{U}$ can be identified with its membership function $A\colon \mathscr{U} \to [0,1]$. Given $\alpha \in [0,1]$ the α-cut of A is given by the set $\{x \in \mathscr{U} \mid A(x) \leq \alpha\}$. Let us recall also the notion of fuzzy partition.

Definition 1. *A fuzzy partition Δ of a universe $\mathscr{U}$ is a set of fuzzy sets $\Delta_1, \ldots, \Delta_n$ on $\mathscr{U}$ fulfilling the covering property, i.e.*

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

for all $x \in \mathscr{U}$ there exists $k \in \{1,\ldots,n\}$ such that $\Delta_k(x) > 0$. The fuzzy sets Δ_k (for $k = 1,\ldots,n$) are called the classes *of* Δ.

In the literature the reader can find several additional condition imposed on fuzzy partitions. As, for instance normality, continuity, convexity or Ruspini condition [22].

Note that the previous definition is generic and $\mathscr{U}$ may be vectors of $\mathbb{R}^n$. Actually, in Section III and later on, the universe $\mathscr{U}$ considered is $\mathbb{R}^n$ for some $n \in \mathbb{N}$, but we keep using the notation $\mathscr{U}$ for the sake of the presentation and to emphasize that such a set is the universe where the fuzzy partition is defined. Moreover, the classes (membership functions) in the fuzzy partitions are defined by combinations of triangular fuzzy sets, which are defined on real numbers (i.e. $\mathscr{U} = \mathbb{R}$) and are determined by 3-tuple of real numbers. In this respect, given $a,b,c \in \mathbb{R}$ satisfying $a \leq b \leq c$, the triangular fuzzy set (a,b,c) is given by the membership function

$$T(a,b,c)(x) = \begin{cases} \frac{x-a}{b-a} & \text{if } a < x \leq b \\ \frac{c-x}{c-b} & \text{if } b < x \leq c \\ 0 & \text{otherwise} \end{cases}$$

Moreover, we do an abuse of notation and allow that $a = -\infty$ or $c = \infty$, in such cases the membership functions of such fuzzy sets are given by:

$$T(-\infty,b,c)(x) = \begin{cases} 1 & \text{if } x \leq b \\ \frac{c-x}{c-b} & \text{if } b < x \leq c \\ 0 & \text{otherwise.} \end{cases}$$

and

$$T(a,b,\infty)(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a < x \leq b \\ 1 & \text{otherwise} \end{cases}$$

respectively.

B. F-transforms

Here we recall the basic notions of (standard) F-transforms (originally introduced in [22]) in the general data framework given in [28]. The main difference between both approaches is that data does not need to have a functional structure, and then, they can be applied to arbitrary datasets. In this respect, we consider a finite subset $\mathbf{T} = \{(x_i,y_i)\}_{i\in\mathbb{I}}$ of $\mathscr{U} \times \mathbb{R}$ without a functional structure; i.e. for the same $x \in \mathscr{U}$ it may exist either two tuples $(x,y_1),(x,y_2) \in \mathbf{T}$ satisfying $y_1 \neq y_2$ or it may exists $x_0 \in \mathscr{U}$ such that $(x_0,y) \notin \mathbf{T}$ for all $y \in \mathbb{R}$. Since $\mathbf{T}$ represents a dataset, we use the following terminology: for simplicity let us assume $\mathscr{U} = \mathbb{R}^n$, each tuple $(x,y) = (x_1,x_2,\ldots,x_n,y) \in \mathbf{T} \subseteq \mathbb{R}^{n+1}$ corresponds to one object b; each coordinate in $\mathbf{T}$ corresponds to a different attribute a_i and it is called variable; each value x_i in $(x_1,x_2,\ldots,x_n,y)$ is the degree of the object b concerning the variable a_i; finally, the variables $(x_1,x_2,\ldots,x_n) \in \mathscr{U}$ are called independent and the variable y is called depended.

Definition 2. *Let $\mathbf{T} = \{(x_i,y_i)\}_{i\in\mathbb{I}} \subseteq \mathscr{U} \times \mathbb{R}$ and let $\Delta = \{\Delta_1,\ldots,\Delta_n\}$ be a fuzzy partition of $\mathscr{U}$. We say that the n-tuple $\mathbf{F}_\Delta[\mathbf{T}] = [F_1,\ldots,F_n] \in \mathbb{R}^n$ is the direct F-transform of $\mathbf{T}$ w.r.t. Δ if*

$$F_k = \frac{\sum_{i\in\mathbb{I}} y_i \cdot \Delta_k(x_i)}{\sum_{i\in\mathbb{I}} \Delta_k(x_i)} \tag{1}$$

for all $k \in \{1,\ldots,n\}$.

It is not difficult to check that definition above extends the original one in the following way: given a function $f\colon \mathscr{U} \to \mathbb{R}$ the Definition 2 coincides with the original definition given in [22] by identifying f with the subset $\mathbf{T}_f = \{(x,f(x) \mid x \in \mathscr{U}\} \subseteq \mathscr{U} \times \mathbb{R}$, i.e., $\mathbf{F}_\Delta[\mathbf{T_f}] = \mathbf{F}_\Delta[f]$. As in the original definition [22], the components of the direct F-transform coincide with a *least squares weighted* solution where the weights are given by the classes of the fuzzy partition Δ chosen, as recalled in the following proposition.

Proposition 1 ([28]). *Let $\mathbf{T} = \{(x_i,y_i)\}_{i\in\mathbb{I}} \subseteq \mathscr{U} \times \mathbb{R}$ and let $\Delta = \{\Delta_1,\ldots,\Delta_n\}$ be a fuzzy partition of $\mathscr{U}$. Then the k-th component of the direct F-transform is the minimum of the following function:*

$$\phi(z) = \sum_{i\in\mathbb{I}} (y_i - z)^2 \cdot \Delta_k(x_i) \tag{2}$$

As in the original approach [22], the inverse F-transform is a function defined from the direct F-transform components.

Definition 3. *Let $\mathbf{T} = \{(x_i,y_i)\}_{i\in\mathbb{I}} \subseteq \mathscr{U} \times \mathbb{R}$ and let $\mathbf{F}_\Delta[\mathbf{T}] = [F_1,\ldots,F_n] \in \mathbb{R}^n$ be the* direct F-transform *of* $\mathbf{T}$ *w.r.t.* Δ. *Then, the function defined, for all $x \in \mathscr{U}$, as:*

$$\mathbf{T}^F_\Delta(x) = \frac{\sum_{k=1}^n F_k \cdot \Delta_k(x)}{\sum_{k=1}^n \Delta_k(x)} \tag{3}$$

is called the inverse F-transform *of* $\mathbf{T}$ *w.r.t.* Δ.

Some remarks about the previous definition. Firstly, the inverse F-transform $\mathbf{T}^F_\Delta(x)$ is a function independently whether the set $\mathbf{T}$ has the structure of a function or not. Secondly, note that the domain of the inverse F-transform is $\mathscr{U}$, so it is defined even for those $x \in \mathscr{U}$ such that there is not $(x,y) \in \mathbf{T}$; i.e., it can be used easily as an interpolation and regression tool. Finally, the inverse F-transform is closely related to the function obtained by assigning to each $x \in \mathscr{U}$ the mean among all the y_i such that $(x,y_i) \in \mathbf{T}$ (see [28] for more details).

C. L_1-F-transforms and QF-transforms

Taking as a reference Proposition 1, [24] proposes a modification of the (standard) direct F-transform as a minimizer of a residual absolute error instead of a residual square error.

Definition 4 ([24][1]). *Let $\mathbf{T} = \{(x_i,y_i)\}_{i\in\mathbb{I}} \subseteq \mathscr{U} \times \mathbb{R}$ and let $\Delta = \{\Delta_1,\ldots,\Delta_n\}$ be a fuzzy partition of $\mathscr{U}$. We say that the n-tuple $\mathbf{F}^{L_1}_\Delta[\mathbf{T}] = [F_1,\ldots,F_n] \in \mathbb{R}^n$ is the L_1-F-transform of $\mathbf{T}$ w.r.t. Δ, if for each $k \in \{1,\ldots,n\}$, F_k is a minimizer[2] of the following function:*

$$\phi(z) = \sum_{i\in\mathbb{I}} |y_i - z| \cdot \Delta_k(x_i) \tag{4}$$

[1]For presentation purposes, the definition is presented directly in the context of non functional data.

[2]The minimizer of a function $f\colon \mathbb{R} \to \mathbb{R}$ is the value $z \in \mathbb{R}$ such that $f(z)$ is the minimum of f.

Here it is worth recalling that the median of a dataset $Y = \{y_i\}_{i\in\mathbb{I}}$ can be characterized as the minimizer of the function $\phi(z) = \sum_{i\in\mathbb{I}}|y_i - z|$. Note that such an expression is similar to Equation (4) but this later is weighted by the classes Δ_i of a fuzzy partition. Then, we can assert that the main difference between the L_1-F-transforms and standard F-transform is that the former is related to the median whereas the latter to the mean. Going one step further, and taking into account that the q-th quantile of a dataset $Y = \{y_i\}_{i\in\mathbb{I}}$ coincides with the minimizer of the objective function:

$$\Phi(z) = \sum_{i\in\mathbb{I}} w_q(y_i)\cdot|y_i - z|,$$

where $w_q(y_i)$ is the weighted function

$$w_q(y_i) = \begin{cases} 1-q & \text{if } y_i < z \\ q & \text{if } y_i \geq z, \end{cases}$$

we can define the direct q-th quantile F-transform (or the q-th QF-transform) as follows.

Definition 5. *Let $\mathbf{T} = \{(x_i, y_i)\}_{i\in\mathbb{I}} \subseteq \mathscr{U}\times\mathbb{R}$, let $\Delta = \{\Delta_1,\dots,\Delta_n\}$ be a fuzzy partition of $\mathscr{U}$ and let $q \in \mathbb{R}$. We say that the n-tuple $\mathbf{QF}^q_\Delta[\mathbf{T}] = [F_1,\dots,F_n] \in \mathbb{R}^n$ is the direct q-th QF-transform of $\mathbf{T}$ w.r.t. Δ, if for each $k \in \{1,\dots,n\}$, F_k is a minimizer of the following function:*

$$\phi(z) = \sum_{i\in\mathbb{I}} w_q(y_i)\cdot|y_i - z|\cdot\Delta_k(x_i) \tag{5}$$

where $w_k(y_i)$ is the weighted function

$$w_q(y_i) = \begin{cases} 1-q & \textit{if } y_i < z \\ q & \textit{if } y_i \geq z. \end{cases}$$

We can interpret the k-th component of the direct q-th QF-transforms as the q-th quantile over the variable y_i by restricting $\mathbf{T}$ to the elements in the class Δ_k. Note also that the definition above is cyclic, but it is well-defined (see [24]). It is worth mentioning also that Definition 5 differs slightly from the original approach in the following way: in [24] the QF-transform is defined as a vector of fuzzy sets (or fuzzy numbers) whereas Definition 5 define QF-transforms as scalar values in $\mathbb{R}$; one per each $q \in [0,1]$. Nevertheless, it is easy to prove that both definitions are equivalent in the sense that one can be obtained from the other and viceversa. This modification is motivated to allow the definition of different fuzzy sets for the consequent of rules during the construction of the knowledge data base (see Section III). Finally, note that each component of the q-th QF-transform can be obtained by means of a weighted linear programming that is straightforwardly constructed [24]. In this way, the computation of QF-transforms can be efficiently performed in practice.

As the inverse standard F-transform, the inverse q-th QF-transform is a function from $\mathscr{U}$ to $\mathbb{R}$.

Definition 6. *Let $\mathbf{T} = \{(x_i, y_i)\}_{i\in\mathbb{I}} \subseteq \mathscr{U}\times\mathbb{R}$, let $q \in [0,1]$ and let $\mathbf{QF}^q_\Delta[\mathbf{T}] = [F_1,\dots,F_n] \in \mathbb{R}^n$ be the direct q-th QF-transform of $\mathbf{T}$ w.r.t. Δ. Then, the function defined for all $x \in \mathscr{U}$ as*

$$\mathbf{T}^{QF}_\Delta(x) = \frac{\sum_{k=1}^n F_k\Delta_k(x)}{\sum_{k=1}^n \Delta_k(x)} \tag{6}$$

is called the q-th inverse *QF*-transform *of* $\mathbf{T}$ *w.r.t.* Δ.

For each $x \in \mathscr{U}$, $\mathbf{T}^{QF}_\Delta(x)$ approximates the q-th quantile of the set $D_x = \{y_i \mid (x, y_i) \in \mathbf{T}\}$. Note that the approximation given by $\mathbf{T}^{QF}_\Delta(x)$ takes into account the elements in $\mathbf{T}$ which first component is *"close"*[3] to x; as a result, it is also applicable when $D_x = \emptyset$.

III. Quantile F-transforms for the definition of fuzzy inference rules

In general, every fuzzy inference system has a similar structure, which consists in the following three steps:

- Fuzzification: this process takes the raw data and converts it into fuzzy data that can be processed by means of fuzzy tools. In this step, the consideration of a suitable fuzzy partitions is fundamental for a good behaviour of the inference system.
- Inference engine: this is the central stage, and consists in the computation of an inference from the input data previously fuzzified. Such an inference requires the use of a knowledge database composed by rules If-Then. The knowledge database used in the inference engine is a crucial for the procedure; it may be either constructed from data or given by experts.
- Defuzification: this is the oposite process to Fuzzification, and takes the outputs of the inference engine, given usually in terms of fuzzy sets, and computes an output in the crisp setting.

Although the Fuzzification process is a fundamental step, we do not go in deep and assume a fixed fuzzy partition Δ. In Section III-A we show how to construct fuzzy association rules by means of direct QF-transforms. Then, we define the Inference engine in Section III-B by using the inverse QF-transform. Finally in Section III-C, we present a simple defuzzification procedure, although others can be defined according to applicational purposes. Finally in Section III-D we provide three measures to determine the significance of the the obtained association rules. Figure 1 shows a diagram with a description of the proposed fuzzy inference structure.

A. Construction of a knowledge database from QF-transforms

For the sake of a better understanding of the fuzzy inference rule system, it is convenient to present firstly the construction of rules in the knowledge database.

In general, the knowledge database used to perform the inference engine is a set of rules with the form:

$$\text{Rule}: \; \mathit{IF} \; (A_1 \;\text{ and }\; A_2 \;\text{ and }\; \dots \;\text{ and }\; A_n) \;\; \text{THEN} \;\; B \tag{7}$$

where A_i (called antecedent) and B (called consequent) represent fuzzy sets. The meaning of the rule is the following: if all the antecedents $A_1, A_2, \dots A_n$ are satisfied by an input, then the output should satisfy the consequent B as well. For example, the following toy rule of a break control system in a car

$$\text{Rule}: \; \mathit{IF} \; (\text{HighSpeed and CloseCurve}) \;\; \text{THEN} \;\; \mathit{StrongBrake}$$

[3]Where the relationship of closeness is given by the fuzzy partition Δ.

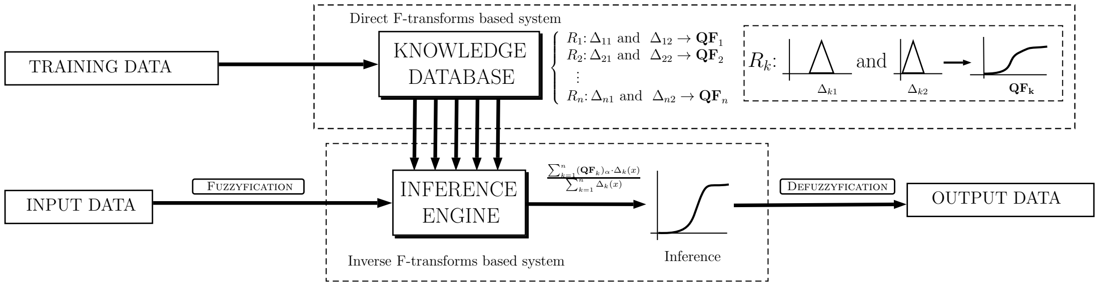

Fig. 1. Diagram of the proposed fuzzy inference system based on quantiles F-transforms.

represents the knowledge *If the car is going fast and is approaching a curve, then the car should break strongly*.

In our approach the construction of the knowledge database is unsupervised, for such a reason we require a training dataset for mining association fuzzy rules. We fix a dataset $\mathbf{T} = \{(x_i, y_i)\}_{i \in \mathbb{I}} \subseteq \mathscr{U} \times \mathbb{R}$, the antecedents of the rules are those classes in a fuzzy partition Δ and the fuzzy sets in the consequents are constructed via direct quantiles F-transforms as follows: for each class $\Delta_k \in \Delta$ we define the fuzzy rule:

$$\text{Rule:} \;\; \textit{IF} \;\; \Delta_k \;\; \text{THEN} \;\; \mathbf{QF}_k \tag{8}$$

where $\mathbf{QF}_k$ is the fuzzy set (on $\mathbb{R}$) defined by the membership function:

$$\mathbf{QF}_k(x) = \min\{q \in [0,1] \mid \left(\mathbf{QF}^q_\Delta[\mathbf{T}]\right)_k \geq x\}$$

for all $x \in \mathbb{R}$; where $\left(\mathbf{QF}^q_\Delta[\mathbf{T}]\right)_k$ is the k-th component of the q-th direct QF-transform whit respect to Δ_k.

Some remarks:

- Although the antecedent is given above by only one class Δ_k, it is worth noting that it may be identified with a conjunction of various fuzzy sets. For example, consider the universe $\mathscr{U} = \mathbb{R}^2$ and two fuzzy partitions for $\mathbb{R}$ given by $\Delta^x = \{\Delta^x_1, \ldots, \Delta^x_n\}$ and $\Delta^y = \{\Delta^y_1, \ldots, \Delta^y_m\}$. Then, we can define a fuzzy partition Δ on $\mathscr{U} = \mathbb{R}^2$ by considering the basic functions defined for all $(x,y) \in \mathbb{R}^2$ by:

  $$\Delta_{ij}(x,y) = \Delta^x_i(x) \cdot \Delta^y_j(y)$$

  for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$. Under the construction of fuzzy rules presented above, there is exactly one rule with antecedent Δ_{ij}. Note that the product between the classes $\Delta^x_i(x)$ and $\Delta^y_j(y)$ can be interpreted as a conjunction. Actually, the product may be replaced by an arbitrary t-norm or fuzzy aggregator function [4]. Therefore, the interpretation of the unique antecedent in the rule of Equation 8 may be considered also a conjunction of fuzzy sets as in Equation (7).
- Note that the consequent of each fuzzy rule, $\mathbf{QF}_k$, is a fuzzy set defined on $\mathbb{R}$. Moreover, thanks to the use of quantiles and its properties related to probability, $\mathbf{QF}_k$ has the following (fuzzy) conditional probabilistic interpretation: given $x \in \mathbb{R}$, $\mathbf{QF}_k(x)$ is the probability that an arbitrary $(x,y) \in \mathbf{T}$ with x in Δ_k satisfies $y \leq \mathbf{QF}_k(x)$. In other words, $\mathbf{QF}_k$ may be considered an approximation of a probability distribution.
- For each class in the fuzzy partition there is a rule even if there is no relation between the antecedents and the consequents. However, this is not a defect since the strength of the inference lies on the fuzzy sets $\mathbf{QF}_k$ constructed in the consequent of the rules. In Section III-D we present a measure to determine the dependence between antecedents and consequents by means of $\mathbf{QF}_k$.

Due to the initial nature of this paper, the knowledge database is constructed in a rudimentary way by considering the whole set of rules definable as in Equation (8). Obviously, this consideration entails an exponential increment of rules according to the number of dependent variables (i.e. dimension of $\mathscr{U}$) and the number of classes in their respective fuzzy partitions. The application of many advanced procedures in the literature can be used to reduce the number of constructed rules; see [10]. However, such a consideration is let to further studies due to the lack of space and, as we already said, the introductory purpose of this approach.

B. Inference engine

Once the construction of the knowledge database has been presented, we can focus on the inference system. Let us recall that the inverse F-transform can be used to approximate the original dependent values of the data used to construct the direct F-transforms (see [25]). Therefore, if we assume that a new piece of data comes from the same context of the training dataset, the obtained fuzzy rules (constructed by means of direct F-transforms) should also describe conveniently its behaviour and then, the use of the inverse F-transform can be used to retrieve the dependent variable.

Following the previous consideration, the inverse F-transform should be the central point of the inference engine. Actually, the inference engine is defined by applying the inverse F-transform on the α-cuts of the consequent fuzzy sets $\mathbf{QF}_k$ in all the rules of our knowledge database (Equation (8)). Formally, given $x \in \mathscr{U}$, the inference of the dependent variable is computed as the fuzzy set $\mathbf{QF}$ whose α-cuts $\mathbf{QF}_\alpha$, with $\alpha \in [0,1]$, is given by:

$$\mathbf{QF}_\alpha = \frac{\sum_{k=1}^{n}(\mathbf{QF}_k)_\alpha \cdot \Delta_k(x)}{\sum_{k=1}^{n}\Delta_k(x)}$$

where $(\mathbf{QF}_k)_\alpha$ denotes the α-cut of the fuzzy set $\mathbf{QF}_k$ associated with the k-th rule in the knowledge base and the (escalar interval) product $\cdot$ is defined by $\lambda \cdot [a,b] = [\lambda a, \lambda b]$, for $\lambda \geq 1$ and $a, b \in \mathbb{R} \cup \{\infty\}$.

Note that this inference reminds the TS-inference system [9] (i.e. the output is computed by a weighted mean) but with two important differences: firstly it reports a fuzzy set instead of a real number and secondly, the weighted mean is used to create α-cuts instead of truth values.

C. Defuzzification

Since the output of the inference engine is a fuzzy set, it is necessary to consider a defuzzification procedure. Let us recall that, since the consequents of the association rules are defined by quantile F-transforms, the inferred fuzzy set **QF** represents the probability distribution of data according to the information acquired from the antecedents; in Section IV we show experimentally that assumption. Therefore, the consideration of the centroid of **QF** as the defuzzification procedure may be meaningless.

One possible defuzzification procedure can be given in terms of intervals. In other words, since the fuzzy set obtained as output represents the distribution of data, we can determine an expected range of the dependent variable. For example, we can assume that the $\alpha\%$ of the data satisfying the antecedents belong to the α-cut $\mathbf{QF}_\alpha$[4]. If we prefer a bounded interval as output, another option is to consider the interval obtained by the intersection of the $\frac{\alpha}{2}$-cut and $1-\frac{\alpha}{2}$-cut, since we can also assume that it approximately contents the $\alpha\%$ of the data satisfying the antecedents.

Many other defuzzification procedures may be defined. For example, we can put effort in determining the kind of distribution (Normal, Exponential, Poisson, etc) of the data and then proceed accordingly; e.g., to determine accumulation of data by the variation of the output. However, the analysis of the different kinds of defuzzification procedures is out of the scope of this paper and it will be studied in future works.

D. Measures of significance

Although the elimination of rules is not considered in this approach, it is important to determine whether one rule is significant or not. By such a reason we propose here three different measures. The first one measures the number of objects in the training dataset that support the obtained rules. Since all the rules are constructed by direct F-transforms and, formally the process consider all the objects in the dataset, the support is given by the number of object that has a significant impact in the definition of $\mathbf{QF}_k$, For such a reason, the support of a rule:

$$R_k\colon\ IF\ \Delta_k\ \text{ THEN }\ \mathbf{QF}_k$$

is defined as the fuzzy cardinality of the antecedent Δ, that is:

$$\text{supp}(R_k) = Card(\Delta_k) = \sum_{u\in\mathscr{U}} \Delta_k(u). \tag{9}$$

[4]Note that the α-cut $\mathbf{QF}_\alpha$ coincides with an interval of the form $[a,\infty]$ for certain $a \in \mathbb{R}$

On the other hand, since we construct an association rule for all class in Δ, we must determine whether there is a dependence between the class $\Delta_k \in \Delta$ and the consequent. Let us recall that the fuzzy set $\mathbf{QF}_k$, in the rule $\Delta_k \to \mathbf{QF}_k$, represents the probability distribution of the variable associated to the consequent *"conditioned to the satisfiability of the antecedent* Δ_k*"*. That is, $\mathbf{QF}_k$ approximates the probability distribution of $\{y_i \mid (x_i,y_i) \in \mathbf{T}$ and $x_i \in \Delta_k\}$. Therefore, we can assume that if $\mathbf{QF}_k$ coincides with the *"non conditioned"* probability distribution of the variable associated to the consequent (i.e, with the probability distribution of $\{y_i \mid (x_i,y_i) \in \mathbf{T}\}$), then there is full independence between variables. Based on such an idea, we propose here the following two measures of dependence between the antecedent and the consequent of a rule $R_k\colon \Delta_k \to \mathbf{QF}_k$ as:

$$\text{dep}_{max} = \max_{x\in[min,max]} |\mathbf{QF_k}(x) - F(x)| \tag{10}$$

$$\text{dep}_{sum} = \frac{1}{max - min} \int_{min}^{max} |\mathbf{QF_k}(x) - F(x)| dx \tag{11}$$

where *min* (resp *max*) is the minimum (resp. maximum) value in the dataset corresponding to the dependent variable and F denotes the probability distribution of the the dependent variable; i.e, of $\{y_i \mid (x_i,y_i) \in \mathbf{T}\}$. Note that both measures are in $[0,1]$ and the closer the measures to 0, the more independence between antecedents and consequent.

IV. Experimental validation

In this section we validate experimentally the fuzzy inference system described in the previous section. For such a reason, we divide the section into two parts. In the first one, we consider synthetic datasets in order to show the behaviour of the fuzzy inference systems under certain circumstances. Subsequently, we apply the approach to real data by considering the dataset uci-combined-cycle-power-plant dataset, available in https://archive.ics.uci.edu, aimed at the estimation of the energy production of a power station. The quantiles F-transforms are discretized by computing 99 quantiles uniformly distributed.

A. Application to synthetic databases.

The goal of this section is to show the profits of the proposed approach in order to acquire knowledge from a training database. For such a reason, we create three groups of datasets for different experiments to analyze the significance of the obtained rules. The first group of datasets is oriented to the analysis of independent variables, the second one oriented to the analysis of functional data and the third one to the analysis of dependent (but not functional) variables. For the sake of a better understanding, we create only datasets of two dimensions.

In the first experiment, we consider two variables A and B and construct a dataset of 500 entries generated randomly by two uniform distributions with range values in $[0,10]$. Then, we consider a uniform fuzzy partition on the variable A of 5 classes and we compute the respective 5 rules in the knowledge database (one per each class in the fuzzy partition) for the

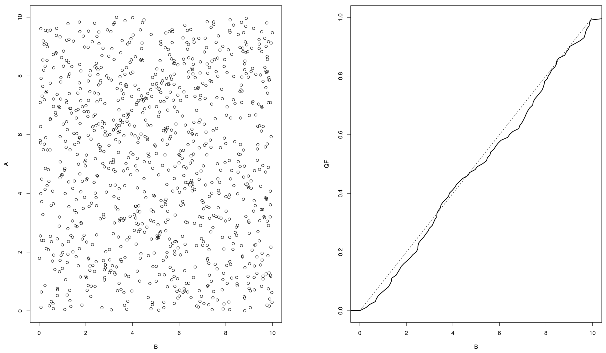

Fig. 2. Left, point cloud of two independent variables. Right, one consequent fuzzy set $\mathbf{QF_k}$, obtained from the independent variable given in the left, and overlapped with the uniform distribution $\mathrm{U}(0,10)$.

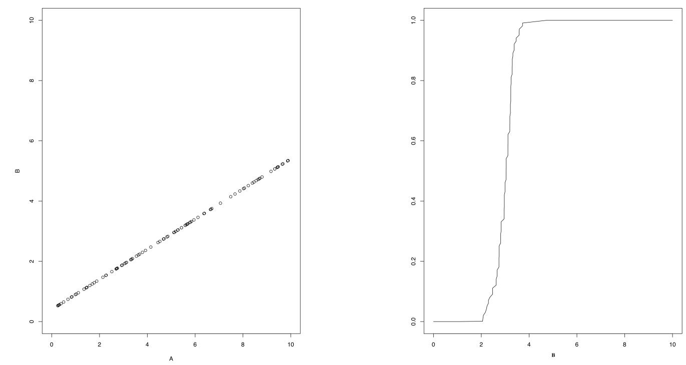

Fig. 3. Left, point cloud of two functional variables under the relation $2y = x + 0.8$. Right, the consequent fuzzy set $\mathbf{QF_3}$ obtained for the triangular class $\Delta_3 = T(3.14, 5.07, 6.99)$; i.e. the rule $\Delta_3 \to \mathbf{QF_3}$.

inference of B: i.e., we compute the respective fuzzy sets $\mathbf{QF}_1, \ldots, \mathbf{QF}_5$ obtained from the direct quantile F-transforms.

In order to evaluate our results, we compute the measures of dependence dep_{max} and dep_{sum} of each rule with respect to the uniform distribution $\mathrm{U}(0,10)$. The experiment has been repeated 20 times (i.e., in total we have analyzed 100 rules) in order to obtain more robust conclusions. The mean (μ) of the measures dep_{max} and dep_{sum} have been $\mu(\mathrm{dep}_{max}) = 0.15$ and $\mu(\mathrm{dep}_{sum}) = 0.086$ with quasi-variance 0.003 and 0.007 respectively. From the previous measures, we can conclude that the measures dep_{max} and dep_{sum} can be used to determine the independence dependence between the variables A and B; as expected. In Figure 2 we show one of the consequent fuzzy sets $\mathbf{QF}_k$ overlapped with the distribution of the uniform distribution to make more visible the similarity between them.

In the next experiment we consider functional data. That is, we consider two variables A and B, and one function $f\colon \mathbb{R} \to \mathbb{R}$ such that $B(x) = f(A(x))$ for every object $u \in \mathscr{U}$. We construct four groups of datasets with 100 objects. In all of them the values of A are (randomly) uniformly distributed between 0 and 10, and $B(x) = f(A(x))$ for different functions $f\colon \mathbb{R} \to \mathbb{R}$. In the first group, B is always constant, i.e., $f(x) = c$ for certain $c \in [0,10]$; in the second one , f is a random straight line; in the third one, f is a second order polynomial; and in the fourth group f is a logarithmic function. In each experiment we have computed the measures of dependence dep_{max} and dep_{sum} for each rule with respect the uniform distribution $\mathrm{U}(0,10)$. We have repeated 20 times the experiment for each group (i.e., 20 randomly generated functions for each group) and the maximum value, the minimum value, the mean and quasi-variance of the measures dep_{max} and dep_{sum} are given in the following table:

	$cte.$	$mx+n$	ax^2+bx+c	$\log_a(x)$
$\max(\mathrm{dep}_{max})$	0.999	0.769	0.914	0.886
$\max(\mathrm{dep}_{sum})$	1	0.424	0.478	0.466
$\min(\mathrm{dep}_{max})$	0.499	0.329	0.297	0.308
$\min(\mathrm{dep}_{sum})$	0.5	0.179	0.149	0.171
$\mu(\mathrm{dep}_{max})$	0.7458	0.573	0.619	0.615
$\mu(\mathrm{dep}_{sum})$	0.67	0.579	0.289	0.302
$\sigma^2_{n-1}(\mathrm{dep}_{max})$	0.022	0.021	0.04	0.034
$\sigma^2_{n-1}(\mathrm{dep}_{sum})$	0.127	0.007	0.011	0.010

Therefore, if we compare these results with those obtained for independent variables, we can observe that the fuzzy rules computed by quantiles F-transforms are able to capture the dependence between variables. In Figure 3 we show the consequent $\mathbf{QF}$ of one rule obtained for functional data; note the difference between $\mathbf{QF}$ an the uniform partition.

In the last experiment developed with synthetical datasets, we consider the straight line $x = y$ in $\mathbb{R}^2$ modified by some noise. In particular, we generate a dataset of 500 objects where the variables $A(x)$ are uniformly distributed in $[0,10]$. Then, we construct a variable B by the equation $B(x) = A(x) + N(0,1)$, where $N(0,1)$ denotes one value randomly generated by the normal distribution with mean 0 and standard deviation 1. In Figure 4, we show one of the obtained dataset with one of the consequent fuzzy sets obtained for one rule. Note the similarity of $\mathbf{QF}_k$ with the normal distribution $N(cor, 1)$, where cor denotes the core of the class Δ_k; in the case illustrated in Figure 4, $cor = 5.07$. The mean of the obtained measures dep_{max} and dep_{sum} for these experiments are 0.496 and 0.255 respectively; the quasi variance obtained for both measures are 0.016 and 0.006 respectively. Comparing these results with the ones obtained for independent data, we may conclude that the obtained rules obtained by quantile F-transforms capture the relationship between the generated variables in this experiment.

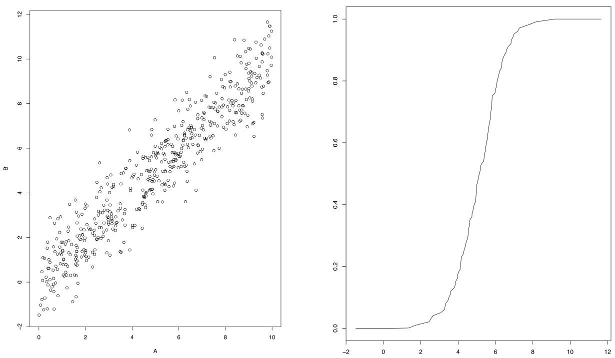

Fig. 4. Left, point cloud of two functional variables under the relation $2y = x + 0.8$. Right the consequent fuzzy set $\mathbf{QF_3}$ obtained from those variables for the triangular class $\Delta_3 = (3.14, 5.07, 6.99)$.

B. Real data

In this section, we consider the uci-combined-cycle-power-plant dataset[5] which was analyzed in [29] under different regression models. The dataset contains 5 hourly average variables, namely: Ambient Temperature (AT), Ambient Pressure (AP), Relative Humidity (RH), Exhaust Vacuum (V) and electrical energy output (EP). The dataset contains 9568 inputs collected from a Combined Cycle Power Plant over 6 years (2006-2011).

Our goal is to illustrate the interpretability of the output as an approximation of a probability distribution. Accordingly, for each $\alpha \in [0,1]$, we determine the percent of tuples $(x_{AT}, x_{AP}, x_{RH}, x_V, x_{EP}) \in \mathbf{T}$ such that the real value of the dependent variable belongs to the α-cut $\mathbf{QF}_\alpha$. For example, if the dependent variable is the electrical energy output (EP), we determine the percent of tuples $(x_{AT}, x_{AP}, x_{RH}, x_V, x_{EP}) \in \mathbf{T}$ such that $x_{EP} \in \mathbf{QF}_\alpha$. Let us recall that according to the definition of the quantiles F-transforms, $\mathbf{QF}$ approximates the probability distribution of the dependent variable and then, each α-cut should contain approximately the $\alpha\%$ of tuples. With this purpose, we perform three experiments. For all them, we split the data into 4 datasets of 2152 objects; one for training and other three for testing. In the first experiment, we consider PE as dependent variable and AT and V as independent variables. For each independent variable we consider a uniform fuzzy partition of 5 classes. Therefore, we obtain 25 (5×5) rules of the type $(AT \text{ and } V) \to PE$. Then, for each dataset (including the training one), we compute the respective inference for each object and determine whether the real value of the variable PE (the real one) belongs to the α-cut of the obtained inference for $\alpha \in \{0.1, 0.2, \ldots, 0.8, 0.9\}$. Figure 5 shows the percent of objects for which its variable PE belongs to the respective α-cut of the inferred fuzzy set $\mathbf{QF}$. The second experiment is similar to the first one, but considering AT as dependent variable and, AP and RH as independent variables. The results are displayed in Figure 6.

[5]available in https://archive.ics.uci.edu.

$\alpha - cut$	D_1^{Training}	D_2	D_3	D_4
0.1	0.0566	0.0548	0.0613	0.0613
0.2	0.1398	0.1375	0.1589	0.1305
0.3	0.2551	0.2592	0.2611	0.2379
0.4	0.3717	0.3833	0.381	0.3638
0.5	0.5001	0.5051	0.5013	0.493
0.6	0.6273	0.6245	0.6333	0.6194
0.7	0.7476	0.7416	0.749	0.7425
0.8	0.8675	0.8601	0.8722	0.8605
0.9	0.9567	0.9456	0.9581	0.9512

Fig. 5. Percent of objects which real value for the variable PE belongs to the respective α-cut of the fuzzy set obtained in the fuzzy inference that uses 25 rules (5×5) of the type $(AT \text{ and } V) \to PE$.

$\alpha - cut$	D_1^{Training}	D_2	D_3	D_4
0.1	0.0882	0.098	0.0975	0.1013
0.2	0.1923	0.1821	0.1807	0.196
0.3	0.2932	0.2806	0.2792	0.289
0.4	0.3986	0.3805	0.387	0.3977
0.5	0.4958	0.4697	0.4735	0.4855
0.6	0.5855	0.5743	0.5645	0.5757
0.7	0.6905	0.7026	0.6798	0.6868
0.8	0.815	0.8238	0.8145	0.8187
0.9	0.9233	0.9154	0.9196	0.9117

Fig. 6. Percent of objects which real value for the variable AT belongs to the respective α-cut of the fuzzy set obtained in the fuzzy inference that uses 25 rules (5×5) of the type $(AP \text{ and } RH) \to AT$.

In the third experiment we followed a similar procedure but considering three independent variables (AT,AP and RH) an as dependent variable to PE. In this latter case we considered fuzzy partitions of 5 classes in each independent variable, but only 111 rules of the 125 $(5 \times 5 \times 5)$ potential rules had support measure (supp) different from 0; that is, only 111 rules are suitable for use. The results are displayed in Figure 7. Note that the values on the three tables are clearly correlated, actually, the Pearson correlation coefficient is greater than 0.99 in all cases. Therefore, we can conclude that the fuzzy sets $\mathbf{QF}$ obtained as the output of the fuzzy inference can be interpreted as an approximation of a conditional probability distribution.

$\alpha - cut$	D_1^{Training}	D_2	D_3	D_4
0.1	0.0497	0.0497	0.0539	0.0576
0.2	0.131	0.1277	0.1491	0.1301
0.3	0.229	0.2332	0.2374	0.2197
0.4	0.3554	0.3605	0.3643	0.3368
0.5	0.4916	0.5041	0.4879	0.479
0.6	0.6259	0.6361	0.6231	0.618
0.7	0.756	0.763	0.7657	0.7448
0.8	0.8926	0.8847	0.8903	0.8763
0.9	0.9628	0.9577	0.9646	0.9577

Fig. 7. Percent of objects which real value for the variable AT belongs to the respective α-cut of the fuzzy set obtained in the fuzzy inference that uses 125 rules $(5 \times 5 \times 5)$ of the type $(AT \text{ and } AP \text{ and } RH) \to PE$.

V. Conclusions and future works

In this paper, we have presented a method to construct fuzzy association rules by means of direct quantile F-transforms. Moreover, we have use this set of rules in a fuzzy inference system where its inference engine is based on the inverse quantile F-transforms. The procedure reminds the Takagi-Sugeno inference but reporting a fuzzy set as the output of the inference procedure. Finally, we have shown in a series of experiments, that the set of association rules constructed by quantiles F-transforms captures the dependencies between the variables and that the output of the fuzzy inference system is related to a probability distribution.

Due to the initial character of this approach, there are different lines of future works. For example, the construction of the knowledge database must be improved by putting effort in the definition of the initial fuzzy partition, by combining the creation of rules with decision trees or by reducing the number of rules. The experimental analysis should be extended in order to encourage the probabilistic interpretation of the inference. Finally, we aim at the application of the proposed fuzzy inference system in practical problems; for example in a classification of anomalous data and/or in the analysis of digital forensics data. For that last goals, we will need to investigate also different defuzzication procedures.

Acknowledgment

Partially supported by the Spanish Ministry of Sciences project PGC2018-095869-B-I00, by the Junta de Andaluca project UMA2018-FEDERJA-001 (European Regional Development Funds) and by the European Cooperation in Science & Technology (COST) Action CA17124.

References

[1] J. F. Silva and S. F. Pinto, "Advanced control of switching power converters," in *Power Electronics Handbook (Third Edition)*, third edition ed., M. H. Rashid, Ed. Boston: Butterworth-Heinemann, ch. 36, pp. 1037 – 1113, 2011.

[2] Y. W. Kerk, K. M. Tay, and C. P. Lim, "Monotone interval fuzzy inference systems," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 11, pp. 2255–2264, 2019.

[3] J. Won and F. Karray, "Toward necessity of parametric conditions for monotonic fuzzy systems," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 2, pp. 465–468, 2014.

[4] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*. Kluwer Academic Publishers, 1999.

[5] G. B. Stamou and S. G. Tzafestas, "Fuzzy relation equations and fuzzy inference systems: an inside approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 6, pp. 694–702, 1999.

[6] P. Vojtáš, "Fuzzy logic programming," *Fuzzy sets and systems*, vol. 124, no. 3, pp. 361–370, 2001.

[7] N. Madrid and M. Ojeda-Aciego, "Measuring inconsistency in fuzzy answer set semantics," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 4, pp. 605–622, 2011.

[8] E. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1 – 13, 1975.

[9] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 1, pp. 116–132, 1985.

[10] S. Guillaume, "Designing fuzzy inference systems from data: An interpretability-oriented review," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 3, pp. 426–443, 2001.

[11] F. Herrera and M. Lozano, *Fuzzy Evolutionary Algorithms and Genetic Fuzzy Systems: A Positive Collaboration between Evolutionary Algorithms and Fuzzy Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 83–130.

[12] N. K. Kasabov, "Learning fuzzy rules through neural networks," in *Proceedings 1993 The First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems*, pp. 137–139, 1993.

[13] M. F. Zarandi, S. Soltanzadeh, A. Mohammadi, and O. Castillo, "Designing a general type-2 fuzzy expert system for diagnosis of depression," *Applied Soft Computing*, vol. 80, pp. 329 – 341, 2019.

[14] E. G. Mansoori, "Frbc: A fuzzy rule-based clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 960–971, 2011.

[15] G. Tsekouras, H. Sarimveis, E. Kavakli, and G. Bafas, "A hierarchical fuzzy-clustering approach to fuzzy modeling," *Fuzzy Sets and Systems*, vol. 150, no. 2, pp. 245 – 266, 2005.

[16] K. Nozaki, H. Ishibuchi, and H. Tanaka, "A simple but powerful heuristic method for generating fuzzy rules from numerical data," *Fuzzy Sets and Systems*, vol. 86, no. 3, pp. 251 – 270, 1997.

[17] L. Zhang, J. Li, and H. Yang, "T–s fuzzy model identification with sparse bayesian techniques," *Neural Processing Letters*, vol. 50, no. 3, pp. 2945–2962, 2019.

[18] R. Koenker, *Quantile Regression*, ser. Econometric Society Monographs. Cambridge University Press, 2005.

[19] S. Gudder, "Fuzzy probability theory," *Demonstratio Mathematica*, vol. 31, no. 1, pp. 235–254, 1998.

[20] L. A. Lucas, T. M. Centeno, and M. R. Delgado, "General type-2 fuzzy inference systems: Analysis, design and computational aspects," in *2007 IEEE International Fuzzy Systems Conference*, pp. 1–6, 2007.

[21] P. Hájek and V. Olej, "Adaptive intuitionistic fuzzy inference systems of takagi-sugeno type for regression problems," in *Artificial Intelligence Applications and Innovations*, L. Iliadis, I. Maglogiannis, and H. Papadopoulos, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 206–216, 2012.

[22] I. Perfilieva, "Fuzzy transforms: Theory and applications," *Fuzzy Sets and Systems*, vol. 157, no. 8, pp. 993 – 1023, 2006.

[23] N. Madrid, "An extension of f-transforms to more general data: potential applications," *Soft Computing*, vol. 21, no. 13, pp. 3551–3565, 2017.

[24] M. L. Guerra, L. Sorini, and L. Stefanini, "Quantile and expectile smoothing based on l1-norm and l2-norm fuzzy transforms," *International Journal of Approximate Reasoning*, vol. 107, pp. 17 – 43, 2019.

[25] I. Perfilieva, V. Novák, and A. Dvořák, "Fuzzy transform in the analysis of data," *International Journal of Approximate Reasoning*, vol. 48, no. 1, pp. 36 – 46, 2008

[26] M. Štěpnička, A. Dvořák, V. Pavliska, and L. Vavříčková, "A linguistic approach to time series modeling with the help of f-transform," *Fuzzy Sets and Systems*, vol. 180, no. 1, pp. 164 – 184, 2011

[27] F. Di Martino and S. Sessa, "Detection of fuzzy association rules by fuzzy transforms," *Advances in Fuzzy Systems*, pp. 1–12, 2012.

[28] N. Madrid, "An extension of f-transforms to more general data: potential applications," *Soft Computing*, vol. 21, no. 13, pp. 3551–3565, 2017.

[29] P. Türecki, "Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods," *International Journal of Electrical Power & Energy Systems*, vol. 60, pp. 126 – 140, 2014.