Application of prediction models using fuzzy sets: an Bayesian inspired approach

Felipo Bacani*, Laécio C. de Barros

Department of Applied Mathematics, State University of Campinas (IMECC – UNICAMP), 13083-970, Campinas–SP, Brazil

Abstract

A fuzzy inference framework based on fuzzy relations is explored and applied to a real set of simulated forecasts and experimental data referring to temperature and humidity in specific coffee crop sites in Brazil. In short, the used model consists of fuzzy relations over possibility distributions, resulting in a fuzzy model analog to a Bayesian inference process. The application of the fuzzy model to temperature and humidity data resulted in a set of revised forecasts, which were later compared to the correspondent set of experimental data using two different statistical measures of accuracy, MAPE (mean absolute percentage error) and Willmott D. Statistical results were confronted to the original simulated forecast fit to experimental data, showing that the methodology was, in most cases, able to improve the specialist's forecasts in both statistical measures.

Keywords: Possibility theory, Fuzzy relations, Fuzzy inference systems

1. Introduction

This work intends to apply a fuzzy inference pattern based on reference [1] to the particular problem of *forecast processing*: suppose a decision-maker having to make a choice based on variable X. In order to do so, a specialist gives him a(n) (imperfect) forecast Y of X and he has at hand a historical record of observed values of X and its corresponding forecasts Y. The decision-maker has to combine the forecast he received with the historical information, taking into account what happened to previous forecasts in the same problem (specialist's skill).

Forecast processing problems have been thoroughly studied under the probabilistic framework, where a solution based on the Bayes formula (Section 2.1.1)) has been proposed by Krzysztofowicz [21–25], in the particular context of hydrology and management of water resources. In this case, forecast skill is modeled by the likelihood function $f(y|x)$. The likelihood function and the priori probability density distribution of X, h0 (x) are determined by the historical data. Krzysztofowicz called his probabilistic framework as *Bayesian Processor of Forecasts* (BPF) and, in reference to BPF, the framework fuzzy used in this work, purposed by Lapointe and Bobée [1], was called *Possibilistic Processor of Forecasts* (PPF).

The PPF framework has some advantages over its Bayesian equivalent as dealing with analytic expressions instead of the usual approximations in the Bayesian context. The fuzzy inference framework consists, in short, of obtaining solutions from a fuzzy relations problem that contains information from historical data. The analogy with the Bayesian process is made by identifying terms analog to a priori distribution, a likelihood function and a posteriori distribution on the fuzzy relations problem described.

2. Fuzzy inference framework

This work has a different, more synthetic exposition of the methodology if compared to the article [1]. The considerations and explorations are restricted here to the product and minimum t-norms and supreme t-conorm.

*Corresponding author

Email addresses: bacani@ime.unicamp.br (Felipo Bacani), laeciocb@ime.unicamp.br (Laécio C. de Barros)

2.1. Preliminaries

2.1.1. Bayesian Inference

The use of the Bayesian methodology in the forecasting context is frequent when historic information for the problem is scarce or useless [14, p. 241]. Let y be a random variable with density function f, characterized by an unknown parameter x. Its density is denoted as $f(y|x)$ to indicate the dependence on the value of x. The probability density function of the parameter x is denoted as $h_0(x)$ and called *priori distribution*. The priori distribution measures the subjective information ("belief degree") about the value of x.

In a forecast situation, suppose that an initial estimate of x is given as a probability distribution $h_0(x)$ (*priori distribution*), and subsequent information about the situation is obtained through random variable y, whose *likelihood function* $f(y|x)$ depends on x. The new estimate of x is obtained in the form of a revised distribution, $h_1(x|y)$, known as *posteriori distribution*. If both random variables x and y are continuous, by Bayes Theorem:

$$h_1(x|y) = \frac{h_0(x)\, f(y|x)}{\int_x h_0(x)\, f(y|x)\, dx}. \tag{2.1}$$

The posteriori distribution $h_1(x|y)$ can be seen as a *tradeoff* between information of the likelihood $f(y|x)$, which represents the evidence of the data by itself, and posteriori distribution $h_0(x)$, which represents external evidences suggestive of higher frequencies.

2.1.2. Fuzzy relations and the Compositional Rule of Inference

T-norms. Triangular norms (t-norms) are used to define fuzzy set aggregation operators. A t-norm is a function $\triangle : [0,1]_\times[0,1] \to [0,1]$, that satisfies $\forall a, a', b, b', c \in [0,1]$:

i) $a \triangle b = b \triangle a$ (commutativity);
ii) $a \leq a',\ b \leq b' \Longrightarrow a \triangle b \leq a' \triangle b'$ (monotonicity);
iii) $(a \triangle b) \triangle c = a \triangle (b \triangle c)$ (associativity);
iv) $a \triangle 1 = a$ (neutral element).

• *Example:* $a \triangle b = a \cdot b = ab$ (product)
• *Example:* $a \triangle b = a \wedge b = \min\{a, b\}$ (mininum)

Fuzzy relations. A fuzzy relation A in the universe $\mathbf{U}\times\,\mathbf{V}$ is any fuzzy subset of $\mathbf{U}\times\,\mathbf{V}$. The inverse relation A^{-1} is a fuzzy relation in $\mathbf{V}\times\,\mathbf{U}$ such that $\forall u \in U, v \in V,\ \varphi_{A^{-1}}(v,u) = \varphi_A(u,v)$. Given a *t-norm* $\triangle$ and the fuzzy relations A in $\mathbf{U}\times\,\mathbf{V}$ and R in $\mathbf{V}\times\,\mathbf{W}$, is denoted as $A \otimes^t R$ the fuzzy relation in $\mathbf{U}\times\,\mathbf{W}$ with pertinence function

$$\varphi_{A\otimes^t R}(u,w) = \sup_{v\in\mathbf{V}} \{\varphi_A(u,v) \triangle \varphi_R(v,w)\}. \tag{2.2}$$

The relation (2.2) is called *sup-t composition* between the fuzzy relations A and R. When the t-norm is the minimum operation, $\triangle = \wedge$, the resulting composition is called $sup - min$ and denoted as $A \circ R$. In particular, if A is a fuzzy subset of $\mathbf{U}$ and R is a fuzzy relation in $\mathbf{U}\times\,\mathbf{V}$, the fuzzy *Compositional Rule of Inference* provides a fuzzy subset B^* of $\mathbf{V}$, denoted as $B^* = A \otimes^t R$ or $B^* = R(A)$, with its pertinence function as a particular case of (2.2):

$$\varphi_{B^*}(v) = \sup_{u\in\mathbf{U}} \{\varphi_A(u) \triangle \varphi_R(u,v)\}. \tag{2.3}$$

This way, in the fuzzy Compositional Rule of Inference, R acts like a functional that takes a fuzzy subset A of $\mathbf{U}$ into the fuzzy subset $B^* = R(A)$ of $\mathbf{V}$.

2.1.3. Fuzzy relational equation

Suppose fuzzy relations A and B, respectively in universes $\mathbf{U} \times \mathbf{V}$ and $\mathbf{U} \times \mathbf{W}$. The fuzzy relation R in $\mathbf{V} \times \mathbf{W}$, maximal (least specific) solution for the problem

$$A \otimes^t R = B, \tag{2.4}$$

is given by[9–11]:

$$\varphi_R(v,w) = \inf_{u\in\mathbf{U}} \{\varphi_A(u,v) \Rightarrow \varphi_B(u,w)\}, \tag{2.5}$$

where "⇒" is a residual fuzzy implication, given by

$$(a \Rightarrow b) = \sup_{z\in[0,1]} \{a \vartriangle z \leq b\}. \tag{2.6}$$

Our interest is in the particular case when A and B are fuzzy subsets of $\mathbf{U}$ and $\mathbf{V}$, respectively. In this case, the maximal solution R of problem (2.2) is simpler than (2.5):

$$\varphi_R(u,v) = (\varphi_A(u) \Rightarrow \varphi_B(v)) = \sup_{z\in[0,1]} \{\varphi_A(u) \vartriangle z \leq \varphi_B(v)\}, \tag{2.7}$$

for an arbitrary t-norm $\vartriangle$.

The fact that in this work R is always given by an implication (2.7), the typical interpretation of a conditional operation is allowed: $\varphi_R(u,v)$ informs the pertinence $\varphi_B(v)$ (of v to B) *given* that u is in A with pertinence $\varphi_A(u)$. This leads to the notation $\varphi_R(u,v) = \varphi_{B|A}(v|u)$. That is, for a given u in A with pertinence $\varphi_A(u)$, $\varphi_R(u,\cdot) = \varphi_{B|A}(\cdot|u)$ is a function in v, which represents the degree of pertinence of v in B, given that u is in A with pertinence $\varphi_A(u)$.

In this notation, the fuzzy Compositional Rule of Inference (2.3) becomes

$$\varphi_{B^*}(v) = \sup_{u\in\mathbf{U}} \{\varphi_A(u) \vartriangle \varphi_{B|A}(v|u)\}, \tag{2.8}$$

with $\varphi_{B|A}(v|u)$ always given by formula (2.7). In (2.8), B^* is not necessarily B[9].

Thus, given A and B, we have R from (2.7) and B^* from (2.8). Now consider S a fuzzy relation in $\mathbf{V}\times\mathbf{U}$ that satisfies the following relational equation $\forall v \in \mathbf{V}$:

$$\sup_{u\in\mathbf{U}} \{\varphi_{B^*}(v) \vartriangle \varphi_S(v,u)\} = \varphi_A(u) \vartriangle \varphi_{B|A}(v|u),^{1} \qquad \forall u \in \mathbf{U}. \tag{2.9}$$

The fuzzy relation S can be obtained from equation (2.9) in a similar way that R was obtained from (2.3). The fuzzy relation S is then given by an expression analog to (2.7),

$$\varphi_S(v,u) = (\varphi_{B^*}(v) \Rightarrow (\varphi_A(u) \vartriangle \varphi_{B|A}(v|u))\ ) = \sup_{z\in[0,1]} \{(\varphi_A(u) \vartriangle \varphi_{B|A}(v|u)) \vartriangle z \leq \varphi_B(v)\}. \tag{2.10}$$

Then, by (2.10), fuzzy relation S can be interpreted as a conditional operation on $v \in \mathbf{V}$, which informs the pertinence of "$\varphi_A(u) \vartriangle \varphi_{B|A}(v|u)$", given the pertinence of v to B^*.

Remark. The adoption of the "sup-t" ($\otimes^t$) operation on the left hand side of (2.9) meets concepts of max-min theories, reflecting "central measures" like fuzzy expected values[12, 13].

2.1.4. Possibility distributions

The model used in this work represents fuzzy sets as possibility distributions, due to the fact that it can be more directly compared to the probability theory than fuzzy sets. This is because possibilities and probability theories work with functions associated with sets that quantify the uncertainty of events [6]. Possibility distributions can be seen as a representation of the pertinence function φ_X of a fuzzy set X [2]. The possibility distribution of a variable X, which takes values on a universe set U, is therefore a function from U to $[0,1]$, denoted by $\pi_X(\cdot)$. π_X is *normal* if there is x^* such that $\pi_X(x^*) = 1$. Unlike the probabilistic case, it is not required that $\int_X \pi_X(x)dx = 1$ for possibility distributions.

[1] In general, the t-norms in the left and right hand sides of equation (2.9) do not need to be the same.

2.2. *Possibilistic Bayes's law*

Using the results of Section 2.1.3 and substituting pertinence functions by possibility distributions for $A \sim X$ and $B \sim Y$, the fuzzy relational problem (2.4) becomes

$$X \otimes^t R = Y, \tag{2.11}$$

and the fuzzy relation R for the problem (2.11) is a rewritten form of (2.7):

$$\pi_R(x, y) = \pi_\triangle(y|x) = \sup_{z\in[0,1]} \{\pi(x) \triangle z \leq \pi_Y(y)\}, \tag{2.12}$$

noticing that in (2.12) the fuzzy relation R has its possibility distribution $\pi_\triangle(y|x)$ as a function of $x \in \mathbf{X}$. Similarly, the fuzzy relation S (in $\mathbf{Y} \times \mathbf{X}$) is a conditional implication for the problem (2.11), but it depends on $y \in \mathbf{Y}$ instead. Therefore $\pi_S(y, x) = \pi_\triangle(x|y)$ is given by a rewritten form of (2.10):

$$\pi_\triangle(x|y) = (\pi^*_\triangle(y) \Rightarrow (\pi(x) \triangle \pi_\triangle(y|x))\ ) = \sup_{z\in[0,1]} \{(\pi(x) \triangle \pi(y|x)) \triangle z \leq \pi^*_\triangle(y)\} \tag{2.13}$$

where $(a \Rightarrow b)$ is the residual implication given by (2.6), and $\pi^*(y) = \pi^*_\triangle(y)$ (2.8) is rewritten as:

$$\pi^*_\triangle(y) = \sup_{x\in\mathbf{X}} \{\pi(x) \triangle \pi_\triangle(y|x)\}. \tag{2.14}$$

The expression (2.13) will be further explored for the two norms of interest in this work, product and minimum.

2.2.1. *Case t-norm=product* $(\cdot)$

For $\triangle = (\cdot)$, (2.6) becomes

$$(a \Rightarrow b) = \begin{cases} 1 & \text{if } a < b \\ \frac{b}{a} & \text{if } a \geq b \end{cases}, \tag{2.15}$$

and as

$$\pi^*_{(\cdot)}(y) = \sup_{x\in\mathbf{X}} \{\pi(x)\, \pi_{(\cdot)}(y|x)\} = a \geq b = \pi_{(\cdot)}(y|x)\, \pi(x),$$

by (2.15), the expression (2.13) is given by

$$\pi_{(\cdot)}(x|y) = \frac{\pi(x)\, \pi_{(\cdot)}(y|x)}{\sup_{x\in\mathbf{X}} \{\pi(x)\, \pi_{(\cdot)}(y|x)\}}. \tag{2.16}$$

The possibility distribution $\pi_{(\cdot)}(x|y)$ (2.16) is clearly similar to the Bayesian posteriori distribution $h_1(x|y)$ (2.1),

$$h_1(x|y) = \frac{h_0(x)\, f(y|x)}{\int_x h_0(x)\, f(y|x)\, dx},$$

where $h_0(x)$ is priori probability distribution and $f(y|x)$ the likelihood function (Section 2.1.1).

The analogy between the terms $\pi_\triangle(x|y) \sim h_1(x|y)$ (posteriori distribution), $\pi(x) \sim h_0(x)$ (priori distribution) and $\pi_\triangle(y|x) \sim f(y|x)$ (likelihood function) must be emphasized on both inference patterns. Interpretations about the priori/posteriori distributions, the likelihood function and about the Bayesian inference process itself are the same in the fuzzy and Bayesian frameworks.

2.2.2. *Case t-norm=minimum* $(\wedge)$

From (2.13), it is possible to verify that if $\triangle = \wedge$, the posteriori possibility distribution is

$$\pi_\wedge(x|y) = \begin{cases} 1 & \text{if } \pi^*_\wedge(y) = \pi(x) \wedge \pi_\wedge(y|x) \\ \pi(x) \wedge \pi_\wedge(y|x) & \text{if } \pi^*_\wedge(y) > \pi(x) \wedge \pi_\wedge(y|x) \end{cases}. \tag{2.17}$$

2.2.3. Comments

i) Expression (2.13) can be seen as the Bayes's law (posteriori distribution) of the PPF inference framework (Section 1), for an arbitrary t-norm. It provides information about the variable X for an given forecast $Y = y$, by combining information of the priori distribution and the likelihood function (Section 2.1.1).

ii) If $\pi^*_\triangle(y) \geq \pi(x),\ \forall x \in \mathbf{X}, y \in \mathbf{Y}$ (Y is less specific than X), by (2.7) the likelihood function always equals 1,

$$\pi_\triangle(y|x) = \sup_{z \in [0,1]} \{\pi(x) \triangle z \leq \pi^*_\triangle(y)\} \equiv 1,$$

by the property (iv) of a t-norm (Section 2.1.2). Substituting this likelihood function in the posteriori expressions for the product (2.16) and minimum (2.17) t-norms (provided the priori distribution is normal), one is able to conclude that the posteriori equals the priori distribution. This means that variable Y is not capable of providing information about variable X, a result observed in the Bayesian case as well.

2.2.4. About the likelihood function

One possible way to obtain the likelihood function in this work is to use, given $\pi(x)$ and $\pi^*_\triangle(y)$, the expression (2.7). Another way to proceed is to obtain the likelihood function directly from the data, that is, without using (2.7). The second alternative was chosen due to the use of the PPF framework in this work. Therefore, the likelihood will be obtained from the historical record. The procedure of doing so is described in Section 2.4.1.

There has been recent works taking more theoretical approaches: in the fuzzy/possibilistic case, Coletti and Vantaggi [15] presented an axiomatization of conditional possibility distributions for an arbitrary t-norm T (T-conditional possibilities). For the Bayesian case, Min & Czado [16] and Aas *et al.* [17] worked in the direction of constructing multivariate probability distributions (likelihood functions as well) using copulas ("pair-copulae") as building blocks, with applications in finance[2].

2.3. Historical remarks

Works searching for equivalent concepts in fuzzy theory and probability theories started back in the 70's, initiated by Zadeh[2] himself. He defined concepts analog to independent and non-interactive random variables[3], obtaining expressions for a conditional fuzzy set from the marginal and joint ones, arriving at more complicated expressions than the probabilistic case due to the use of a minimum operator instead of a product.

Zadeh's ideas developed further by the works of Nguyen [3], who proposed a particular normalization that guarantees the consistency of the non-interactivity concept defined by Zadeh. Nguyen goes on obtaining expressions for the "conditional" fuzzy set. Hisdal [4] goes on the opposite direction of Zadeh and Nguyen, obtaining the joint distribution from the conditional and marginal ones, and managed to (in a sense) generalize the expressions obtained by Zadeh. In the 80's, Bouchon [5] provided an important reference for the approach that is adopted here, as it was apparently the first work that treated the subject as a fuzzy relational equations problem, which was the direction taken by Lapointe and Bobée.

2.4. An example

In this example, it is intended to obtain information about variable X based on Y, a forecast of X. This inference framework was called Possibilistic Processor of Forecasts (PPF, Section 1). This section aims to apply the PPF's methodology and equations (Section 2.2) to an example set of data (Figure 1). That is, to obtain the priori distribution and likelihood function from the historical data (Section 2.4.1), and use the expressions (2.16,2.17) to obtain the posteriori distribution for the t-norms of product and minimum (Section 2.4.2).

Remark on notation. In PPF framework, the likelihood function and priori distribution are obtained directly from the historical record. Then, only the posteriori distribution $\pi_\triangle(x|y)$ and $\pi^*_\triangle(y)$, expressions (2.13) and (2.14) of Section 2.2, respectively, depend on the t-norm $\triangle$. So the notation for the likelihood function is $\pi(y|x)$ instead of $\pi_\triangle(y|x)$, and it remains $\pi(x)$ for the priori distribution.

---

[2]Copulas are aggregation operators found in the probabilistic theory that played a similar role played by t-norms in the fuzzy context.

[3] Independent and non-interactive random variables are equivalent concepts in probability theory, what is not true for Zadeh's fuzzy case. This is due to Zadeh's replacement of the product operator for the minimum in equations from the probability theory.

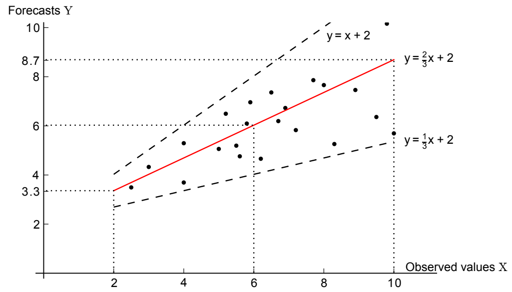


Figure 1: Example of historical data: Observed values X vs. Forecasts Y.

2.4.1. Obtaining priori distribution and likelihood function from example historical record

Priori distribution. Suppose that after analysis of the historical record from Figure 1, the conclusion about the behavior of variable X (independently of forecast Y, for that is the priori information) is that most values of X lies between 5 and 8, and in some special cases, values are smaller than 5, but never below 2; and, if greater than 8, it is never above 10. This information allows to assign the larger pertinence (possibility) "1" to the interval of higher occurrence [5,8], and linearly extends this value to the extremes 2 and 10. This results in a trapezoidal fuzzy number (Figure 2a):

$$\pi(x) = trapez(2,5,8,10)(x) = \begin{cases} \frac{(x-2)}{3} & \text{if } 2 < x < 5 \\ 1 & \text{if } 5 \leq x \leq 8 \\ \frac{(10-x)}{2} & \text{if } 8 < x < 10 \\ 0 & \text{otherwise} \end{cases}. \tag{2.18}$$

Likelihood function. Taking into account that this function represents the *relation* between the observed values X and forecasts Y, this relation can be portrayed by Figure 1. Supposing the relation between X and Y is linear, the linear least squares curve can be used to designate the points (x, y) that have the strongest relationship, that is, they are the points presenting greater possibility ("1").

Figure 1 also suggests that the prediction error increases with X, that is, the distance between the points and the least squares curve increases with X. This observation is corroborated by the fact that the upper dashed line that encompasses the points of the graphic superiorly presents an inclination greater than one; – as the lower dashed line has an inclination smaller than one. Therefore it is reasonable to ask that the *support of the likelihood function should increase with X*. This can be done if one limits the support superiorly by the upper dashed line of the figure and inferiorly by the lower dashed line. A way of translating these observations into a mathematical expression is to model the likelihood function by a triangular fuzzy number "$triang(a, u, b)(y)$", which attains its maximum pertinence value at $y = u$, limited above by $y = b$ and below by $y = a$. That is,

$$triang(a, u, b)(y) = \begin{cases} \frac{(y-a)}{u-a} & \text{if } a \leq y < u \\ \frac{(y-b)}{u-b} & \text{if } u \leq y < b \\ 0 & \text{otherwise} \end{cases}, \tag{2.19}$$

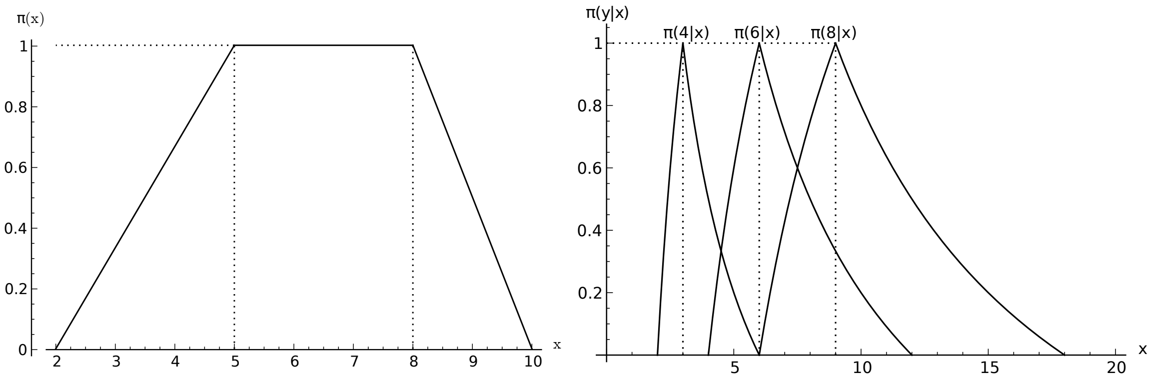
and by substituting "$u = \frac{2}{3}x + 2$", "$b = x + 2$", "$a = \frac{x}{3} + 2$" (from Figure 1) in (2.19) above and isolating x, the following is obtained

$$\pi(y|x) = \begin{cases} 3 - 3\frac{(y-2)}{x} & \text{if } y - 2 \leq x \leq \frac{3}{2}(y-2) \\ 3\frac{(y-2)}{x} - 1 & \text{if } \frac{3}{2}(y-2) \leq x \leq 3(y-2) \\ 0 & \text{otherwise} \end{cases}, \tag{2.20}$$

that is the expression of the likelihood function (Figure 2b) for this example.

2.4.2. Posteriori distribution for forecast $Y = 4$ from example historical record

Suppose the forecast for the variable X is $Y = 4$. To obtain the posteriori distribution of X given the prevision $Y = 4$ it is necessary to use the expressions from Section 2.2, as well the priori and likelihood obtained for the example data (Section 2.4.1).



(a) Priori distribution (2.18).

(b) Likelihood function (2.21) for a few forecast values Y.

Figure 2: Priori distribution (a) and likelihood function (b) for the example data.

Likelihood function (2.21) for $y = 4$, as illustrates Figure 2b, becomes

$$\pi(4|x) = \begin{cases} 3 - \frac{6}{x} & \text{if } 2 \leq x \leq 3 \\ \frac{6}{x} - 1 & \text{if } 3 < x \leq 6 \\ 0 & \text{otherwise} \end{cases} . \tag{2.21}$$

The likelihood function (2.21) represents the values expected for variable X given forecast $Y = 4$. That is, it is more likely that X assumes value 3, with neighboring values possibility distributing non-linearly around 3. Before actually calculating the posteriori distribution, it is possible to discuss a little about its behavior: as the posteriori distribution can be seen as a tradeoff between information of the likelihood function and the priori distribution (Section 2.1.1), and the value with higher possibility for the likelihood has little relevance for the priori, $\pi(x = 3) = 1/3$ by (2.18), it is fair to infer that the posteriori will take this unbalance into account and assign as its maximum a value with a higher priori relevance (that is, other than $X = 3$), without sacrificing much of its likelihood relevance[4].

The following are the calculations of posteriori distribution for the t-norms of product and minimum. The following procedure is repeated for both t-norms: first the possibility distribution "$\pi(4|x) \vartriangle \pi(x)$" is obtained, and then the supreme is taken (2.14) to obtain $\pi^*_{\vartriangle}(y = 4)$. The posteriori expression results by applying the two expressions obtained to (2.16) or (2.17), according to the particular t-norm.

Posteriori, t-norm=product $(\cdot)$**.** From (2.18) and (2.21):

$$\pi(x) \cdot \pi(4|x) = \begin{cases} \frac{x^2-4x+4}{x} & \text{if } 2 \leq x \leq 3 \\ \frac{-x^2+8x-12}{3x} & \text{if } 3 \leq x \leq 5 \\ \frac{6}{x} - 1 & \text{if } 5 \leq x \leq 6 \end{cases} . \tag{2.22}$$

Evaluating (2.14) for (2.22):

$$\pi^*_{(\cdot)}(y = 4) = \sup_x \{\pi(x)\pi(4|x)\} = \max\left(\sup_{2\leq x\leq 3} \left\{\frac{x^2-4x+4}{x}\right\}, \sup_{3\leq x\leq 5} \left\{\frac{-x+8x-12}{3x}\right\}, \sup_{5\leq x\leq 6} \left\{\frac{6}{x} - 1\right\} \right) = \max\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{5}\right) = \frac{1}{3} \tag{2.23}$$

Substituting (2.22) and (2.23) in (2.16), the following posteriori distribution $\pi^{(\cdot)}(x|4)$ is obtained:

$$\pi_{(\cdot)}(x|4) = \begin{cases} \frac{3(x^2-4x+4)}{x} & \text{if } 2 \leq x < 3 \\ \frac{-x^2+8x-12}{x} & \text{if } 3 \leq x < 5 \\ \frac{18}{x} - 3 & \text{if } 5 \leq x < 6 \end{cases} , \tag{2.24}$$

with its graphic shown in Figure 3a.

[4]It should be noticed that this discussion is independent of the particular t-norm. Actually the t-norm dictates the tradeoff tolerance.

Posteriori, t-norm=minimum ($\wedge$). The possibility distribution "$\pi(4|x) \wedge \pi(x)$" demands the evaluation of the minimum between the priori $\pi(x)$ and the likelihood $\pi(4|x)$ for every x in the domain. This can be done by finding the roots $\{\hat{x}_i\}$ of the equation "$\pi(x) - \pi(4|x) = 0$", and then finding the minimum for each interval between the roots $\{\hat{x}_i\}$. For the equations (2.18) and (2.21), $\hat{x} = 3.77$ and the result is an piecewise function:

$$\pi(4|x) \wedge \pi(x) = \begin{cases} \frac{x-2}{3} & \text{if } 2 \leq x \leq 3.77 \\ \frac{6}{x} - 1 & \text{if } 3.77 \leq x \leq 6 \\ 0 & \text{otherwise} \end{cases} \tag{2.25}$$

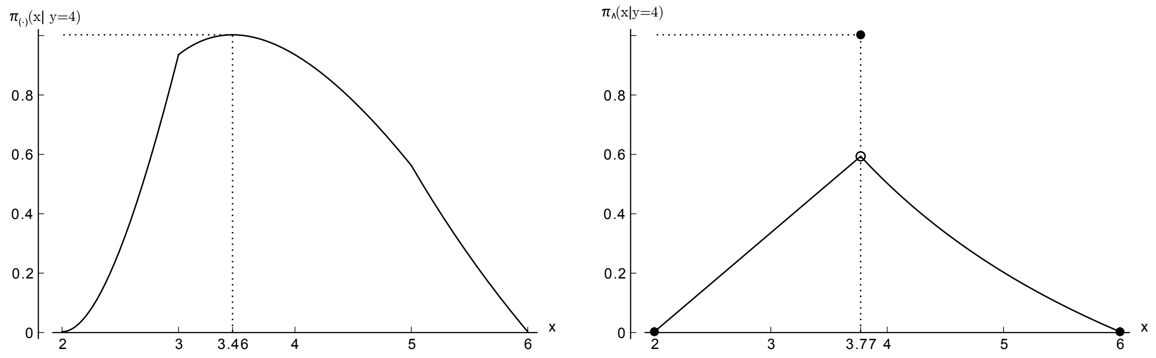
Evaluating (2.14) for (2.25):

$$\pi^*_\wedge(y=4) = \sup_x \{\pi(4|x) \wedge \pi(x)\} = \max\left(\sup_{2\leq x\leq 3.77} \left\{\frac{x-2}{3}\right\}, \sup_{3.77\leq x\leq 6} \left\{\frac{6}{x} - 1\right\}\right) = \max\left(\frac{1.77}{3}, \frac{6}{3.77} - 1\right) = \frac{6}{3.77} - 1 \tag{2.26}$$

and by (2.25) and (2.26), $\pi(4|\hat{x}) \wedge \pi(\hat{x}) = \frac{6}{3.77} - 1 = \pi^*_\wedge(y=4)$. Then the posteriori (2.17) is:

$$\pi_\wedge(x|4) = \begin{cases} \frac{x-2}{3} & \text{if } 2 \leq x < 3.77 \\ 1 & \text{if } x = 3.77 \\ \frac{6}{x} - 1 & \text{if } 3.77 < x \leq 6 \\ 0 & \text{otherwise} \end{cases}, \tag{2.27}$$

illustrated in Figure 3b.



(a) Posteriori, t-norm=product, $\pi_{(\cdot)}(x|4)$ (2.24).

(b) Posteriori, t-norm=minimum, $\pi_\wedge(x|4)$ (2.27).

Figure 3: Posteriori distributions for forecast $Y = 4$, with priori distribution (2.18) and likelihood function (2.21) obtained from example historical data of Figure 1, for product and minimum t-norms.

Comments on posteriori distributions of the example. After simple considerations, at the beginning of the Section 2.4.2, it seemed fair to state that the most relevant value of the resulting posteriori distribution was not $X = 3$. And it is indeed true, as the maximum values of the posteriori distributions are, for both t-norms, different from $X = 3$: $\hat{x} = 3.46$ for the product t-norm and $\hat{x} = 3.77$ for the minimum (Figure 3).

3. Application

The fuzzy inference framework purposed in Section 2.2 is applied to a set of data from coffee crop sites located in Southern Minas Gerais, Brazil. The collected and simulated data refers to the work of Priscila Coltri[28], and aims to explore different consorts between coffee and other plant species in order to obtain microclimates with lower temperatures. Each micro-climate has meteorological stations capable of collecting temperature and humidity information hourly. A forecast is associated with each of these measurements, being obtained by the computational simulation of the corresponding micro-climate using the software ENVI-met [5]. Therefore, given one of the possible micro-climates, every day of data collections informs 24 pairs "measure/forecast" for temperature and 24 pairs for humidity.

The objective is to apply PPF's (Section 1) fuzzy inference framework to obtain, for every measurement day, a set of (24) revised forecasts expected to have better fit to the experimental data than the original (specialist's) forecasts. As the PPF framework is used both in the example of Section 2.4 and in the application, the example will serve as a guideline here, with some points treated in a higher level of detail. The notation here is the same of the example.

3.1. About the data / Test methodology

Each set of data corresponds to three days of collections on a same micro-climate in a given season. There are four micro-climates, and data was collected in the summer and in the winter. Therefore, the temperature variable has eight data sets, as humidity has seven (one was discarded). One known test methodology is to separate the set of data into two parts: training group and testing group (see [9, p. 268]). That is, the priori distribution and the likelihood function are obtained using the data from the training group, and they are used to obtain the posteriori distribution for the forecasts that belong to the testing group.

Three tests were made upon the separation of the testing and training groups of each data set (corresponding to three days) in a 66/33% proportion: 1) the first of the three days is the training group, the last two days are the testing group; 2) second day as the training group, the first and third days are the testing group; and 3) the third day is the training group, the first two days are the testing group. This kind of data separation is made in order to have a greater variability for the tests. In this application 36 tests were made, of which 24 are for temperature and 12 for humidity.

3.2. Two choices for the priori distribution

From the example of Section 2.4, one forecast ($Y = 4$) originated *two* posteriori distributions, one for product and one for minimum t-norm. For this application, besides working with the two t-norms it is worked with *two different priori distributions*: a trapezoidal one (like the example of Section 2.4) and a normalized parabola. This way, each forecast originates *four* posteriori distributions, each one corresponding to a priori/t-norm combination.

3.3. Performing the tests

For a given data set consisting of three days, one day is chosen as the testing group. Therefore, the other two days will be the training group. The 48 measure/forecast pairs of the training group are used to obtain the two priori distributions and likelihood function (Section 3.4), and the resulting distributions are used to calculate the posteriori distributions for each forecast of the testing group.

That is, let $\{(x_i, y_i), i = 1, ..., 24\}$ be the testing group, and for a particular forecast y_{i^*}, denote $\hat{x}^1_{i^*}$, $\hat{x}^2_{i^*}$, $\hat{x}^3_{i^*}$, $\hat{x}^4_{i^*}$ as the values with higher possibility of each one of the four resulting posteriori distributions (each one corresponding to a priori/t-norm combination) associated with this forecast [6]. Having this procedure repeated for each forecast y_i ($i = 1, ..., 24$), all the necessary information for this test is obtained:

$$\left\{\mathbf{x}, \mathbf{y}, \hat{\mathbf{x}}^1, \hat{\mathbf{x}}^2, \hat{\mathbf{x}}^3, \hat{\mathbf{x}}^4\right\} = \left\{\left(x_i,\ y_i,\ \hat{x}^1_i,\ \hat{x}^2_i,\ \hat{x}^3_i,\ \hat{x}^4_i\right), i = 1, ..., 24\right\}$$

.

3.4. Obtaining the priori distribution and likelihood functions from historical data

Likelihood function. As pointed out in Section 2.2.4 and in the example of Section 2.4, the likelihood function will be obtained directly from the data. An important assumption of the example in is that the historical data (Figure 1) *already contained* the two dashed limit curves, which was actually the *support* of the likelihood function (Section 2.4.1).

The procedure used in this work to obtain the limit curves is simple: it uses the quality of the linear least squares curve (represented as the determination coefficient $R = r^2$) to define the "looseness" that the dashed curves will give to the data. Thus a good fitting will have its dashed curves very close to the farthest points of the graphic, implying a likelihood function with a sharp support.

Systematically, for a least squares fit "$y = Ax + B$", the upper dashed curve $b(x)$ is obtained by the procedure:

[5] *http://www.envi-met.com/*

[6] If an analog notation were used in the example of Section 2.4, $y = 4$ and $\hat{x}^{\mathbf{1}} = 3.46$ and $\hat{x}^{\mathbf{2}} = 3.77$.

1. One find $(\hat{x}_M, \hat{y}_M)$, the point vertically farthest from the least squares, among all points above the least squares;
2. Find the inclination $\hat{a}_M$ such that "$\hat{y}_M = \hat{a}_M\,\hat{x}_M + B$";
3. The upper dashed curve is obtained by the expression

$$b(x) = (\hat{a}_M + C \cdot |\hat{a}_M - A|)x + B,$$

where $C = (1 - R) + 0.10$, and $0 \leq R \leq 1$ is the determination coefficient of the least squares fit.

The constant C defines the so-called "looseness" of the curve, with a better fit implying a smaller C. The base value 0.10 is a way to force a minimal loose to the curves, and was chosen empirically. An analog procedure is made to find the *lower* dashed curve, given by the expression "$a(x) = (\hat{a}_m - C \cdot |\hat{a}_m - A|)x + B$", with C being the same as before and $\hat{a}_m$ being the inclination that intercepts the farthest of the points below the least squares fit.

Priori distribution. Both the trapezoid and the normalized parabola are obtained from a similar procedure, that is, using a histogram and a cumulative histogram made from the training group data. Only the measure values "x" (the first value of each measure/forecast pair) were used to assemble the histograms, as the priori is concerned only with information on the collected data.

3.5. Statistical analysis

The statistical analysis of a test consists in comparing the forecasts revised by the method to the experimental values, and checking whether their revised forecasts get closer to the experimental values rather than the original forecasts. This means (Section 3.3) to check if the columns $\hat{\mathbf{x}}^j$ ($j = 1, ..., 4$) approaches $\mathbf{x}$ better than $\mathbf{y}$.

This comparison can be exemplified through a graphic like the Figure 4, which compares a particular set of revised forecasts (say, $\hat{\mathbf{x}}^4$) to the set of specialist's forecasts $\mathbf{y}$ and to the experimental values $\mathbf{x}$.

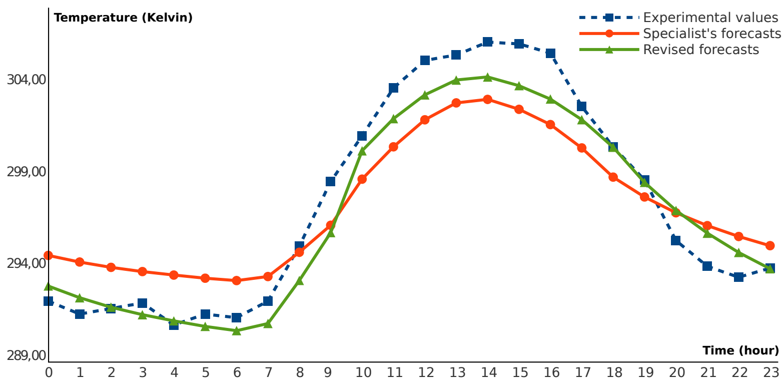


Figure 4: Resulting information from a specific test, with corresponding specialist's forecasts and experimental values.

For this case, it can be concluded that, in the first eight hours, the revised forecasts fitted better experimental data than the specialist's forecasts. Then there is a spike, and next a decay of temperature, both followed better by the revised forecasts over the specialist ones. Although graphics as Figure 4 can illustrate well the behavior of the method, they are neither a reliable (nor a practical) way of drawing conclusions about the method performance. Statistical indicators (shown below) are more suitable for this work.

Statistical indicators. The two ways chosen to compare measure/forecast vectors were: MAPE (Mean absolute percentage error)[27] and Willmott $\mathcal{D}$ [26]. MAPE was chosen for its recognition, and Willmott $\mathcal{D}$ because it was the indicator the specialist already used. If $\mathbf{a}$ and $\mathbf{f}$ are the respective vectors of actual values and corresponding forecasts, their MAPE and $\mathcal{D}$ values (respectively $I^{\mathcal{M}}$ and $I^{\mathcal{D}}$) are given by the expressions

$$I^{\mathcal{M}}(\mathbf{a},\mathbf{f}) = \frac{100\%}{n}\sum_{i}^{n}\left|1 - \frac{f_i}{a_i}\right|, \qquad I^{\mathcal{D}}(\mathbf{a},\mathbf{f}) = 1 - \frac{\sum_{i=1}^{n}|f_i - a_i|^2}{\sum_{i=1}^{n}(|f_i - \bar{a}| + |a_i - \bar{a}|)^2}, \tag{3.1}$$

where $\bar{a}$ is the average value of vector $\mathbf{a}$. Notice that, for all $\mathbf{a}, \mathbf{f}$, $0 \leq I^{\mathcal{D}}(\mathbf{a}, \mathbf{f}) \leq 1$ and $0 \leq I^{\mathcal{M}}(\mathbf{a}, \mathbf{f}) < \infty$. For Willmott $\mathcal{D}$, the value "0" corresponds to the worst set of forecast possible, and "1" to a set of perfect forecasts. As for MAPE, perfect forecasts have a "0" value, and there is no upper bounds for bad forecasts[7].

Criteria for tests validation. Willmott $\mathcal{D}$ between the experimental values $\mathbf{x}$ and its j^{th} set of revised forecasts $\hat{\mathbf{x}}^{\mathrm{j}}$ is denoted as $I^{\mathcal{D}}(\mathbf{x}, \hat{\mathbf{x}}^{\mathrm{j}}) \equiv I_j^{\mathcal{D}}$ ($j = 1, 2, 3, 4$), and between experimental values and the original forecasts $\mathbf{y}$ is $I^{\mathcal{D}}(\mathbf{x}, \mathbf{y}) \equiv I_{\text{orig}}^{\mathcal{D}}$. The same notation $I_j^{\mathcal{M}} \equiv I^{\mathcal{M}}(\mathbf{x}, \hat{\mathbf{x}}^{\mathrm{j}})$ ($j = 1, 2, 3, 4$) and $I_{\text{orig}}^{\mathcal{M}} \equiv I^{\mathcal{M}}(\mathbf{x}, \mathbf{y})$ holds for MAPE indicator. The statistical analysis is made using these indicators to check if the forecasts revised by the method are closer to the experimental values than original forecasts, by calculating the percentage improvements $P_j^{\mathcal{D}}$ and $P_j^{\mathcal{M}}$:

$$P_j^{\mathcal{D}} = \frac{I_j^{\mathcal{D}} - I_{\text{orig}}^{\mathcal{D}}}{I_{\text{orig}}^{\mathcal{D}}}, \qquad P_j^{\mathcal{M}} = \frac{I_{\text{orig}}^{\mathcal{M}} - I_j^{\mathcal{M}}}{I_{\text{orig}}^{\mathcal{M}}}, \qquad j = 1, 2, 3, 4. \tag{3.2}$$

The statistical analysis is made using the values $P_j^{\mathcal{D}}$, $P_j^{\mathcal{M}}$ for $j = 1, 2, 3, 4$. For example, if there is a j^* such that $P_{j^*}^{\mathcal{M}} > 0$ and $P_{j^*}^{\mathcal{D}} > 0$, this means that *at least one* of the four sets of revised forecasts improves the original ones. This is chosen as the "first criterion" used to determine if a test is well succeeded. A "second criterion", sharper although much more reliable, is to check if $P_j^{\mathcal{M}} > 0$ and $P_j^{\mathcal{D}} > 0$ for $j = 1, 2, 3, 4$. This means that *all four sets* of revised forecasts improved the original ones.

4. Conclusions

The statistical results of the 36 tests were:

- 32 tests (89%) were validated through the "first criterion";
- 25 tests (69%) were validated through the "second criterion".

Comments on the developed algorithm. In order to implement the PPF fuzzy inference framework to a set of real data, some tasks required automation. Then the development of a software that uses symbolic calculations[8] to evaluate the analytic procedures involved in the calculation of the posteriori distribution was indispensable. Besides, the systematization of a procedure to obtain the fuzzy likelihood function from a given historical data (Section 3.4) allowed it to be also incorporated to the developed software, so calculations on a large scale could be made possible.

Final comments.

- The method used shows potential for more applications in every problem that deals with decision making and, more importantly, this approach is totally independent from the particular way of generating the forecasts.
- According to the stipulated criteria, the particular application of the method showed good performance. Since it was not defined which one of the four t-norm/priori combinations is the most adequate to obtain the posteriori distribution, the "second criterion" gains more relevance. A possible direction for future investigations is to restrict it to special cases (or families) of t-norms.

[7] This is not a problem for this particular application, as only relative (not absolute) MAPE values matters for the analysis. Besides, the denominators of MAPE expressions will never approach zero as temperature is in Kelvin and the humidity did not approached zero in any of the dealt situations.

[8] This was made using SAGE, an open-source mathematical software, *http://www.sagemath.com*.

Acknowledgements

Financial support by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), process number 306872/2009-9. Professor Fernando Gomide for the helpful comments, and professor Jurandir Zullo for having gently provided the data used in this work.

[1] S. Lapointe and B. Bobée. Revision of possibility distributions: A bayesian inference pattern. *Fuzzy Sets and Systems*, 116(2):119–140, 2000.

[2] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1):3–28, 1978.

[3] H. T. Nguyen. On conditional possibility distributions. *Fuzzy Sets and Systems*, 1(4):299–307, 1978.

[4] E. Hisdal. Conditional possibilities independence and noninteraction. *Fuzzy Sets and Systems*, 1(4):283–297, 1978.

[5] B. Bouchon. Fuzzy inferences and conditional possibility distributions. *Fuzzy Sets and Systems*, 23:33–41, 1987.

[6] D. Dubois, S. Moral, and H. Prade. A semantics for possibility theory based on likelihoods. *J. Math. Anal. Appl.*, 205:359–380, 1997.

[7] D. Dubois and H. Prade. Bayesian conditioning in possibility theory. *Fuzzy Sets and Systems*, 92:223–240, 1997.

[8] D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum, New York, 1988.

[9] F. Gomide and W. Pedrycz. *Fuzzy Systems Engineering: Toward Human-Centric Computing*. John Wiley & Sons, 2007.

[10] W. Pedrycz. Fuzzy relational equations with triangular norms and their resolution. *Busefal*, 11:24–32, 1982.

[11] G. J. Klir and B. Yuan. *Fuzzy sets and fuzzy logic : theory and applications*. Upper Saddle River, N.J.: Prentice-Hall, 1 edition, 1995.

[12] M. Sugeno. *Theory of fuzzy integral and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[13] R.C. Bassanezi and L.C. Barros. A simple model of life expectancy with subjective parameters. *Kybernets 7*, pages 91–98, 1995.

[14] D. C. Montgomery. *Forecasting and time series analysis*. St. Louis: McGraw-Hill, New York, 1976.

[15] G. Coletti and B. Vantaggib. T-conditional possibilities: Coherence and inference. *Fuzzy Sets and Systems*, 160:306–324, 2009.

[16] A. Min and C. Czado. T-conditional possibilities: Coherence and inference. *Journal of Financial Econometrics*, 8(4):511–546, 2010.

[17] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44:182–198, 2009.

[18] K.S. Kelly and R. Krzysztofowicz. Probability distributions for flood warning systems. *Water Resources Research*, 30(4):1145–1152, 1994.

[19] K.S. Kelly and R. Krzysztofowicz. Bayesian revision of an arbitrary prior density. In *Proc. Section on Bayesian Statistical Science of the American Statistical Association*, pages 50–53, 1995.

[20] K.S. Kelly and R. Krzysztofowicz. A bayesian model and choice of experts. In *TIMS/ORSA Joint National Meeting, Boston, MA, 24–27 April*, Department of Systems Engineering, University of Virginia, Charlottesville, 1994.

[21] R. Krzysztofowicz. Bayesian models of forecasted time series. *Water Resources Bulletin*, 21(5):805–814, 1985.

[22] R. Krzysztofowicz. Markovian forecast processes. *J. Amer. Statist. Assoc.*, 82(397):31–37, 1987.

[23] R. Krzysztofowicz. A theory of flood warning systems. *Water Resources Research*, 29(12):3981–3994, 1993.

[24] R. Krzysztofowicz and S. Reese. Bayesian analyses of seasonal runoff forecasts. *Stochastic Hydrology and Hydraulics*, 5:295–322, 1991.

[25] R. Krzysztofowicz and L.M. Watada. Stochastic model of seasonal runoff forecasts. *Water Resources Research*, 22(3):296–302, 1986.

[26] C. J. Willmott. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research*, 90(I5):8995–9005, 1985.

[27] J. S. Armstrong and F. Collopy. Error measures for generalizing about forecasting methods: empirical comparisons. *International Journal of Forecasting*, 8:69–80, 1992.

[28] P. P. Coltri. *Mitigação de emissões de gases de efeito estufa e adaptação do café arábica a condições climáticas adversas*. PhD thesis, Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola, Campinas/SP, 2012. Processo: 565.387/2008-1 – Edital: MCT/CNPq n°27/2007. In portuguese.

[29] F. Bacani. Modelos de predição utilizando lógica fuzzy: uma abordagem inspirada na inferência bayesiana. Master's thesis, Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica, Campinas/SP, 2012. In portuguese.