

Automatic Generation of a Type-2 Fuzzy System for Time Series Forecast based on Genetic Programming

Marco Antônio da Cunha Ferreira^a and Ricardo Tanscheit^a and Marley Vellasco^a

^a DEE, PUC-Rio, Rio de Janeiro, RJ, Brasil , {macfe,ricardo,marley}@ele.puc-rio.br

Abstract

This work describes the development of a type 2 Fuzzy Inference System by using Genetic Programming for applications in time series forecasting. The resulting model, called GPFIS-Forecast+, is based on the GPFIS-Forecast created previously, which made use of Multigenic Genetic Programming and provided good results. Results show that the system with type 2 fuzzy sets improves the performance, especially for noisy data.

Keywords: Fuzzy System, Mamdani, Genetic Programming, Forecasting, Noisy Data.

1 Introduction

Forecast models for time series have been applied successfully in areas such as logistics [20], investment management [7], energy charge forecast [28], meteorology, signal processing [25], etc. Such models may consider statistical techniques, assuming that the time series represents a stochastic process. This approach usually makes use of data processing methods to ultimately execute the forecast. Examples can be found in [1, 2, 9, 26]. Other models may be generated from existing data and then used to predict future data. In this process Artificial Neural Networks [27], Genetic Algorithms [4], Fuzzy Logic [25], [23] and hybridizations of such techniques [8] [11, 12] may be employed. One of these hybridizations is the GPFIS-Forecast [15], which, in essence, is a fuzzy-genetic computational model of the Pittsburgh type [8] that makes use of Multigenic Genetic Programming [16, 21] to generate an optimal rule base for a type 1 Mamdani Fuzzy Inference System. Experimental results [5] have attested the good performance of this model.

This paper describes the development of the GPFIS-Forecast+ model, whose structure is also of the Mamdani type but making use of interval type 2 fuzzy

sets [13] in order to obtain a more efficient rule base and to improve the performance, especially in the presence of noisy data. According to [14], the use of an interval type 2 Takagi-Sugeno fuzzy system demonstrates a great advantage in terms of accuracy when compared to a type 1 Takagi-Sugeno fuzzy system. This paper demonstrates that this is also valid for GPFIS-Forecast+ when compared to GPFIS-Forecast.

2 Multigenic Genetic Programming

Genetic Programming (GP) [16] uses the concept of evolving solutions to resolve problems. The population is characterized by individuals – such as computer programs – that represent possible solutions to the proposed problem. These individuals are structured in trees, which depict a combination of operations (orange circle in Figure 1) and terminals (blue circles in Figure 1).

Terminals represent the variables to be solved, while the connections are mathematical functions (operations). The example of Figure 1 shows a chromosome whose representation as a function is $X * X$, that is, multiply the variable X by itself.

As various types of connectors and terminals are added, the number of possible individuals expresses the complexity of solutions to the problem. Genetic programming uses the concepts of genetic algorithms [8] such as selection, crossover, and mutation to find an optimal solution.

Multigenic Genetic Programming (MGPP) [16] considers that an individual has more than one tree in its representation. Thus a solution is composed by the aggregation of trees. The example of Figure 2 illustrates a chromosome with two genes, each with a tree.

This chromosomal structure opens the way for other possibilities of crossing: high-level (Figure 3) and low level (Figure 4).

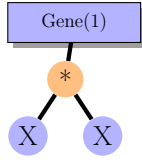


Figure 1: Example of a genetic programming chromosome

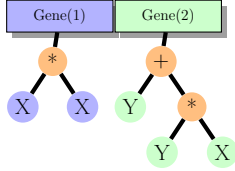


Figure 2: Example of a chromosome as multigenic of genetic programming

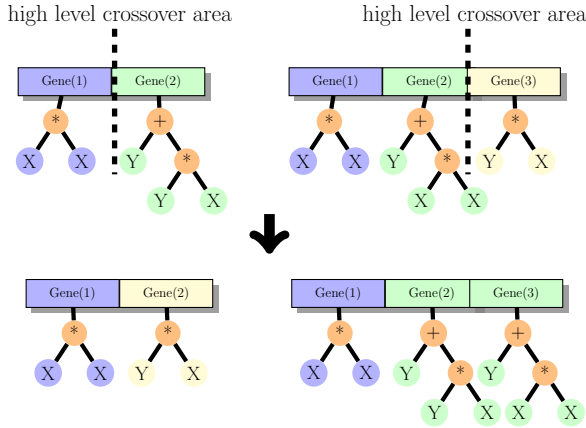


Figure 3: Example of high level crossover

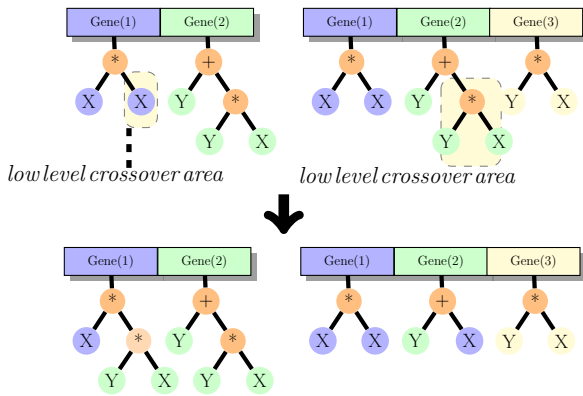


Figure 4: Example of low level crossover

The mutation operator is used to rewrite an entire gene or to change part of a tree of a selected gene.

Individuals in GP are probabilistically selected based

upon a fitness score. Individual programs with the best fitness are more likely than inferior individual programs [22] to provide offspring for the next generation. The selection method used here is the tournament [19].

To preserve the best individuals from a population during the evolution process, elitism replacement [19] is employed. A percentage $n\%$ of the best individuals from the previous population are preserved and mixed to the individuals $1-n\%$ of the offspring; the remaining individuals are deleted. This makes the population size constant during the evolution.

3 GPFIS-Forecast+ model

Time series forecast uses information from the series itself to determine future values. For a time series of size T , the i -th prediction window, whose size is P , is given by $\mathbf{s}_i = \{s_{i,1}, \dots, s_{i,p}\}$, where $i \in [1, (T - P)] \in \mathbb{Z}$. The window \mathbf{s}_i is formed by observations $s_{i,p}$ in each p -th lag S_p , where $p \in [1, P] \in \mathbb{Z}$. The forecast target is denoted by $y_i \in Y$.

GPFIS-Forecast+ is a genetic fuzzy system of the Pittsburgh type [8] which, when adapted to deal with interval type 2 fuzzy sets, assumes the configuration shown in Figure 5 .

The functions of the different blocks are described as follows.

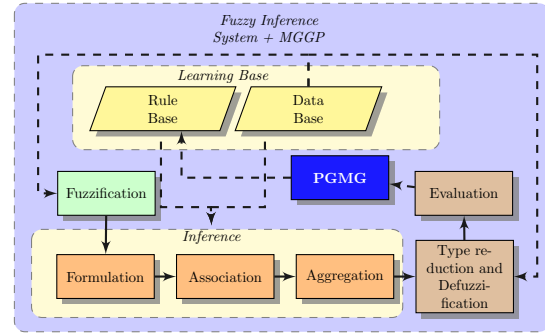


Figure 5: GPFIS - Forecast Type 2

3.1 Fuzzification

Each lag $s_{i,p}$ is associated to L interval type 2 fuzzy sets \tilde{A}_{lp} , whose footprint of uncertainty (FOU) [18] is bounded by sets \overline{A}_{lp} and \underline{A}_{lp} which are, respectively, the upper (UMF) and lower (LMF) membership functions of the interval-type 2 fuzzy set \tilde{A}_{lp} :

$$\overline{A}_{lp} = \left\{ (s_{i,p}, \mu_{\overline{A}_{lp}}(s_{i,p})) | s_{i,p} \in S_p, l \in [1, L] \in \mathbb{Z} \right\} \quad (1)$$

$$\underline{A_{lp}} = \left\{ (s_{i,p}, (1 - FOU) \times \mu_{\bar{A_{lp}}}(s_{i,p})) | s_{i,p} \in S_p, \right. \\ \left. l \in [1, L] \in \mathbb{Z} \right\} \quad (2)$$

The function $\mu_{A_{lp}} : S_p \times [0, 1] \rightarrow [0, 1]$ maps each observation s_p to a degree of relevance to A_{lp} . The Footprint of Uncertainty (FOU), which defines the distance between UMF and LMF, can vary from 0 (equivalent to a fuzzy type 1 system) to 0.9.

In the fuzzification process fuzzy sets of the triangular type [24] are used and the universe is defined by the maximum and minimum values of the training data. Figure 6 shows an interval type 2 fuzzy set, delimited by the upper (red) and lower (blue) fuzzy sets.

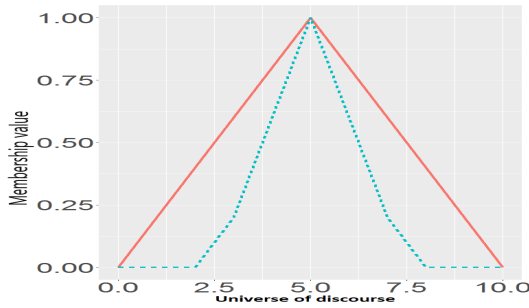


Figure 6: Type 2 fuzzy triangle fuzzy set example.

3.2 Inference

A fuzzy rule is composed of the combination of premises and consequents. An example of a rule with P antecedents and one consequent is:

IF S_1 IS $\tilde{A}_{1,1}$ AND ... AND S_p IS $\tilde{A}_{l,p}$
THEN S_t IS \tilde{B}_k

The premise of this rule can be expressed mathematically by equation 3, where a t-norm operator $*$ represents the “AND” connective. The t-norm used here is the product.

$$\mu_{\tilde{A}_d}(s_{i,P}) = \mu_{\tilde{A}_{1,1}}(s_{i,1}) * \dots * \mu_{\tilde{A}_{l,P}}(s_{i,P}) \quad (3)$$

where $\mu_{\tilde{A}_d}(s_{i,1}, \dots, s_{i,P}) = \mu_{\tilde{A}_d}(s_{i,P})$ is the joint membership degree of the P lags of s_i with respect to the d -th premise ($d \in [1, D] \in \mathbb{Z}$).

From the point of view of Genetic Programming, the premise is a tree and the rule base is a tree chromosome, as shown in Figure 7. The first premise, $d = 1$, for the first windows forecast, $i = 1$, represents, in this case:

$$\mu_{\tilde{A}_1}(s_{1,P}) = \mu_{\tilde{A}_{1,1}}(s_{1,1}) * \mu_{\tilde{A}_{3,2}}(s_{1,2})$$

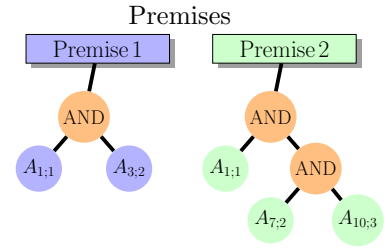


Figure 7: Example of Assumptions of a Rule Base

This rule may be expressed in a linguistic form: “IF S_1 IS $\tilde{A}_{1,1}$ AND S_3 IS $\tilde{A}_{3,3}$ ”.

In the inference stage, the fuzzified data undergo a series of operations according to the rules created in a process consisting of two parts: formulation and association.

In Formulation, for the initial population of genetic programming, premises of the rules of all individuals are created in a random way, taking care that there are no repeated premises for the same individual. As the population evolves, the premises of the new individuals are created through genetic processes such as crossover and mutation.

Association is the process that generates the consequent for each premise created in the formulation stage. This can be done in two ways: (i) randomly or (ii) by using methods that generate consequents from the data. This second option is the one adopted in this work.

The metric that makes it possible to choose the most appropriate consequent for a given premise is the Restricted Minimum Least Squares (MLS):

$$\min : z = \sum_{i=1}^T (\hat{\mu}_{\tilde{B}_k}(y_i) - \sum_{d^{(k)}=1}^{D^{(k)}} w_{d^{(k)}} \times \mu_{\tilde{A}_{d^{(k)}}}(s_{i,P}))^2$$

$$s.t. : \sum_{d=1}^{D^{(k)}} w_{d^{(k)}} = 1 \text{ and } w_{d^{(k)}} \geq 0 \quad (4)$$

where $w_{d^{(k)}}$ weight or influence grade of $\mu_{\tilde{A}_{d^{(k)}}}(s_{i,P})$ for predicting values associated to the k -th consequent term \tilde{B}_k (K fuzzy sets \tilde{B}_k , were $k \in [1, \dots, K] \in \mathbb{Z}$, are associated with Y (output variable)).

The d -th fuzzy rule of a FIS of the Mamdani type can then be expressed as:

$$R_{Mam}^d : \mu_{\tilde{Z}_d}(u) = \mu_{\tilde{A}_d}(s_{i,P}) * \mu_{\tilde{B}_k}(u) \quad (5)$$

where $u \in U$ (universe of discourse of the membership function \tilde{B}_k).

When the rule R_{mam}^d is activated, one output fuzzy set \tilde{Z}_d is generated with a range of memberships for each value of the universe of discourse U . When more than one fuzzy rule out of the D rules are activated at the same time, an aggregation operator is used so that each value of u has only one range of memberships, (one for the UMF and one for the LMF of the activated n rules).

An aggregation operator g (maximum [3], in this work) performs the following operation:

$$\mu_{\tilde{Z}_g}(u) = g(\mu_{\tilde{Z}_1}(u), \dots, \mu_{\tilde{Z}_D}(u)) \quad (6)$$

At the end of the aggregation process the result is an interval type 2 fuzzy set \tilde{Z}_g defined as:

$$\tilde{Z}_g = 1 / [\underline{Z}_g, \overline{Z}_g] \quad (7)$$

Note, in equation 7, that the secondary membership function [18] is equal to one for the entire universe of discourse.

3.3 Type reduction and Defuzzification

The defuzzification in a type 2 FIS must be preceded by a type reduction process. The Karnik-Mendel (KM) [13] method is used here for type reduction, producing the following results for the minimum and maximum values of the centroids interval.

$$\begin{aligned} cl(\tilde{A}) &= \min_{\forall \theta \in [LMF(x), UMF(x)]} \frac{\sum_{i=1}^N x_i \theta}{\sum_{i=1}^N \theta_i} \\ cr(\tilde{A}) &= \max_{\forall \theta \in [LMF(x), UMF(x)]} \frac{\sum_{i=1}^N x_i \theta}{\sum_{i=1}^N \theta_i} \end{aligned} \quad (8)$$

where A_e is the encapsulated fuzzy set of \tilde{A} [18].

Defuzzification is then performed by:

$$Output = \frac{cl(H_L) + cr(H_R)}{2} \quad (9)$$

where H_L and H_R are the values in the universe of discourse that correspond to the approximated minimum and maximum values of the centroids interval.

3.4 Evaluation

The Evaluation procedure in GPFIS-Forecast+ is defined by an error minimization objective. This objective is responsible for the ordering of individuals in the population.

The metric used for comparison of configurations is the Symmetric Mean Absolute Percent Error (SMAPE)

[10]. The well known Mean Squared Error (RMSE) is the metric used for evaluation throughout the experiment, including training and test stages.

4 Case study

The objective of this case study was to investigate how rule bases evolved with the addition of the footprint of uncertainty. To this end, the parameters shown in Table 1 were used in the experiments with GPFIS-Forecast+.

Table 1: GPFIS-Forecast+ configuration parameters

| Parameters | Config. 1 | Config. 2 | Config. 3 |
|----------------------|-----------|------------|------------|
| Forecast repetitions | 10 | 10 | 10 |
| FIS type | 1 | 2 interval | 2 interval |
| Window | 8 | 8 | 8 |
| Forecast horizon | 18 | 18 | 18 |
| Generations | 100 | 100 | 100 |
| RB max. size | 50 | 50 | 50 |
| Tree depth | 3 | 3 | 3 |
| Population size | 49 | 49 | 49 |
| Number of terms | 5 | 5 | 5 |
| FOU Parameter | 0.00 | 0.90 | 0.50 |

The time series Mackey-Glass [17], described by Equation 10, was used to assess the model's performance.

$$\frac{dx(t)}{dt} = a \frac{x(t-\tau)}{1 + x(t-\tau)^n} - bx, \quad a, b, n > 0, \quad (10)$$

with $\tau = 70$; $b = 0.1$; $a = 0.2$; $n = 10$. To solve the equation the following parameters were used: initial condition $x(0) = 1.2$; integration step $\Delta t = 10$, number of samples $\kappa = 145$.

Figure 8 shows the training, validation, and test sets for the series with $\tau = 70$.

For each of the three GPFIS-Forecast+ configurations, 10 predictions were generated. The best prediction for the validation step, for each configuration, was considered in the execution of the forecast with the test base.

5 Results and discussions

Figure 9 shows the predictions for configurations 1 (Forecast T1) and 2 (Forecast T2). Forecast 3 is omitted for the sake of clearness. The columns of Table 2 show the results for GPFIS-Forecast+ using type 1 fuzzy sets, type 2 fuzzy sets with a large uncertainty and type 2 fuzzy sets with intermediate uncertainty, respectively. The results are very similar for the training stage. However, in the test stage, the rule base generated for FIS type 2 with large uncertainty provides a better generalization.

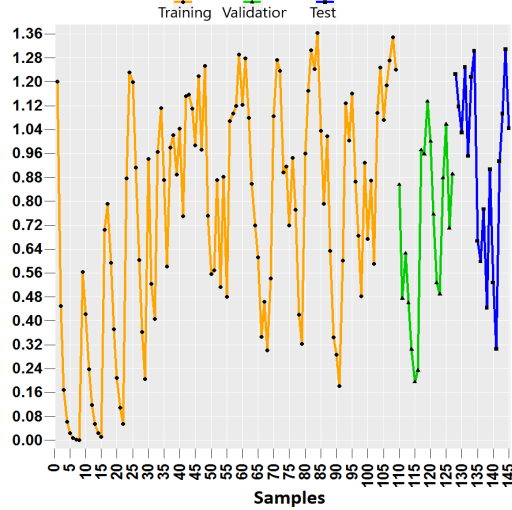


Figure 8: Mackey-Glass Temporary Series

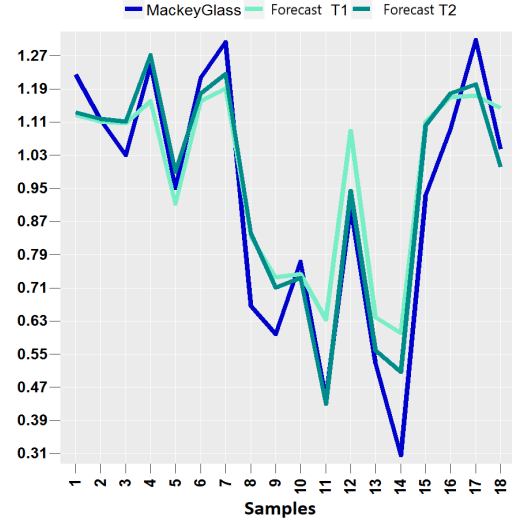


Figure 9: Mackey-Glass time series forecast for $\tau = 70$

Table 2: Results: GPFIS-Forecast+

| Parameters | Config. 1 | Config. 2 | Config. 3 |
|------------------|-----------------|-----------------|-----------------|
| Qty of forecasts | 10 | 10 | 10 |
| Type of FIS | 1 | 2 intervalar | 2 intervalar |
| Window | 8 | 8 | 8 |
| Forecast horizon | 18 | 18 | 18 |
| Generation | 100 | 100 | 100 |
| Max. size BR | 50 | 50 | 50 |
| Tree depth | 3 | 3 | 3 |
| Population size | 49 | 49 | 49 |
| Qty of terms | 5 | 5 | 5 |
| Parameter FOU | 0.00 | 0.90 | 0.50 |
| Results | | | |
| Train SMAPE | 0.171698 | 0.171882 | 0.168916 |
| Train RMSE | 0.088474 | 0.085984 | 0.086360 |
| Validation SMAPE | 0.108982 | 0.203182 | 0.131101 |
| Validation RMSE | 0.067631 | 0.096191 | 0.078356 |
| Test SMAPE | 0.279744 | 0.150133 | 0.279646 |
| Test RMSE | 0.197588 | 0.105722 | 0.195962 |
| BR final size | 10 | 8 | 10 |
| Exec. time (h) | 8 | 9 | 8.1 |

Table 3 presents important characteristics of the model: there are few rules in the final rule base, each rule has a maximum of three antecedents and the number of terms of the linguistic variable is sufficient to generate a good accuracy. The acronyms in the tables mean: “VL” = Very low, “L” = low, “M” = medium, “H” = high and “VH” = very high. The term Lag that succeeded by a number (1 to 8) denotes the element of the forecasting window. It should be noted that a window of size 8 provided the best results for the problem in question.

In the first population, each of the 49 individuals had 25 premises. The Association stage defines which of

these premises can generate valid rules (premises associated with consequents), resulting in the elimination of excessive rules, thus providing more compact rule bases. Note that only lags 2, 3 and 8 appear in the resulting rules.

Table 3: RB of configuration 3

| | Antecedents.1 | Antecedents.2 | Consequents |
|----|---------------|---------------|-------------|
| 1 | L Lag 2 | VL Lag 2 | VL |
| 2 | VL Lag 2 | L Lag 2 | VL |
| 3 | L Lag 2 | L Lag 2 | H |
| 4 | M Lag 2 | L Lag 2 | VH |
| 5 | VL Lag 8 | | VL |
| 6 | L Lag 2 | M Lag 2 | VH |
| 7 | M Lag 2 | M Lag 2 | H |
| 8 | H Lag 2 | H Lag 2 | M |
| 9 | VH Lag 2 | H Lag 2 | L |
| 10 | L Lag 1 | M Lag 2 | VH |

The P-value computed by the Friedman test for the results of the validation step was 0.045049. This shows that there is a statistical difference between results. A paired comparison with the post-hoc Holm [6] test (Table 4) shows that the results for configurations 1 and 2 have no relevant statistical difference; these two stand out when compared to configuration 3.

Table 4: Post Hoc comparison Table for $\alpha = 0.05$ (FRIEDMAN)

| Configuration | $z = (R_0 - R_i)/SE$ | p | Holm |
|------------------|----------------------|----------|-------|
| Config. 2 with 3 | 2.459675 | 0.013906 | 0.025 |
| Config. 2 with 1 | 0.894427 | 0.371093 | 0.05 |

6 Conclusions and Future Work

In this work the GPFIS-Forecast+ model was presented, destined to the forecast of time series. This model showed good accuracy and good interpretability. The use of a type 2 FIS provided excellent results for a noise database when compared to those obtained by the type 1 version.

Future works will consider strategies to initialize the evolution process. An automatic strategy for generating the initial configuration will allow GPFIS-Forecast+ to begin the evolution process with more informative data, possibly making learning easier and faster.

References

- [1] G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, *Time series analysis: forecasting and control*, John Wiley & Sons, 2015.
- [2] R. G. Brown, R. F. Meyer, The fundamental theorem of exponential smoothing, *Operations Research* 9 (5) (1961) 673–685.
- [3] T. Calvo, A. Kolesárová, M. Komorníková, R. Mesiar, Aggregation operators: properties, classes and construction methods, in: T. Calvo, G. Mayor, R. Mesiar (Eds.), *Aggregation Operators*, Vol. 97 of *Studies in Fuzziness and Soft Computing*, Physica-Verlag HD, 2002, pp. 3–104.
- [4] C. T. Chai, C. H. Chuek, D. Mital, T. T. Huat, Time series modelling and forecasting using genetic algorithms, in: *First International Conference on Knowledge-Based Intelligent Electronic Systems.*, Vol. 1, 1997, pp. 260–268.
- [5] M. A. da C. Ferreira, A. S. Koshiyama, M. M. Vellasco, R. Tanscheit, Aprimoramentos de um sistema fuzzy-genético para análise de séries temporais, in: C. J. A. Bastos Filho, A. R. Pozo, H. S. Lopes (Eds.), *Anais do 12º Congresso Brasileiro de Inteligência Computacional*, Curitiba, PR, 2015, pp. 1–6.
- [6] S. García, A. Fernández, J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability, *Soft Computing* 13 (10) (2009) 959.
- [7] P. Hajek, Predicting corporate investment/non-investment grade by using interval-valued fuzzy rule-based systems—a cross-region analysis, *Applied Soft Computing* 62 (2018) 73–85.
URL <http://www.sciencedirect.com/science/article/pii/S1568494617306427>
- [8] F. Herrera, L. Magdalena, Genetic fuzzy systems: A tutorial, *Tatra Mt. Math. Publ.(Slovakia)* 13 (1997) 93–121.
- [9] C. C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, *International Journal of Forecasting* 20 (1) (2004) 5–10.
- [10] I. Ilies, H. Jaeger, O. Kosuchinas, M. Rincon, V. Sakenas, N. Vaskevicius, Stepping forward through echoes of the past: forecasting with echo state networks, Short report on the winning entry to the NN3 financial forecasting competition, available online at http://www.neural-forecasting-competition.com/downloads/NN3/methods/27-NN3_Herbert_Jaeger_report.pdf 53 (2007) 76.
- [11] J.-S. Jang, ANFIS: Adaptive-network-based fuzzy inference system, *IEEE Transactions on Systems, Man, and Cybernetics* 23 (3) (1993) 665–685.
- [12] D. Karaboga, E. Kaya, Adaptive network based fuzzy inference system (anfis) training approaches: a comprehensive survey, *Artificial Intelligence Review* (2018) 1–31.
- [13] N. N. Karnik, J. M. Mendel, Q. Liang, Type-2 fuzzy logic systems, *IEEE transactions on Fuzzy Systems* 7 (6) (1999) 643–658.
- [14] A. Khosravi, S. Nahavandi, D. Creighton, D. Srinivasan, Interval type-2 fuzzy logic systems for load forecasting: A comparative study, *IEEE Transactions on Power Systems* 27 (3) (2012) 1274–1282.
- [15] A. S. Koshiyama, M. M. Vellasco, R. Tanscheit, Development of a fuzzy rule-based system using genetic programming for forecasting problems, in: *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2015, pp. 1–7.
- [16] J. R. Koza, *Genetic programming: on the programming of computers by means of natural selection*, MIT press, 1992.
- [17] M. C. Mackey, L. Glass, et al., Oscillation and chaos in physiological control systems, *Science* 197 (4300) (1977) 287–289.
- [18] J. M. Mendel, R. B. John, Type-2 fuzzy sets made simple, *IEEE Transactions on fuzzy systems* 10 (2) (2002) 117–127.
- [19] R. Poli, W. B. Langdon, N. F. McPhee, J. R. Koza, *A field guide to genetic programming*, Lulu.com, 2008.

- [20] L. Ritzman, L. Krajewski, *Administração da produção e operações*, Pearson Prentice Hall, 2003.
- [21] D. Searson, M. Willis, G. Montague, Co-evolution of non-linear pls model components, *Journal of Chemometrics* 21 (12) (2007) 592–603.
- [22] K. Staats, Genetic programming applied to rfi mitigation in radio astronomy, Ph.D. thesis, University of Cape Town (2016).
- [23] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, in: *Readings in Fuzzy Sets for Intelligent Systems*, Elsevier, 1993, pp. 387–403.
- [24] D. R. Unune, M. M. Barzani, S. S. Mohite, H. S. Mali, Fuzzy logic-based model for predicting material removal rate and average surface roughness of machined nimonic 80a using abrasive-mixed electro-discharge diamond surface grinding, *Neural Computing and Applications* 29 (9) (2018) 647–662.
- [25] L.-X. Wang, J. M. Mendel, Generating fuzzy rules by learning from examples, *IEEE Transactions on Systems, Man, and Cybernetics* 22 (6) (1992) 1414–1427.
- [26] P. R. Winters, Forecasting sales by exponentially weighted moving averages, *Management science* 6 (3) (1960) 324–342.
- [27] G. Zhang, B. E. Patuwo, M. Y. Hu, Forecasting with artificial neural networks:: The state of the art, *International Journal of Forecasting* 14 (1) (1998) 35–62.
- [28] K. Zor, O. Timur, A. Teke, A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting, in: *6th International Youth Conference on Energy (IYCE)*, 2017, pp. 1–7.