Use of Scalable Fuzzy Time Series Methods to predict Electrical Demand

José Miguel Rubio León[1][0000−0003−0377−4397] and José Manuel Rubio Cienfuegos[2]

[1] Universidad Bernardo O'Higgins, Avda Viel 1497, Santiago de Chile, Chile
josemiguel.rubio@ubo.cl
[2] Universidad de Chile Avda Beauchef 850, Santiago de Chile. Chile
jose.rubio@ug.uchile.cl

Abstract. Electric power is one of the main engines of humanity since the late eighteenth century, therefore, understanding the behavior of the demand of this corresponds to a very important study. Within this field, electricity demand prediction has been one of the most developed activities in several researches, where the use of fuzzy time series models has had good results. This paper presents the use of scalable fuzzy time series methods, which were developed by Petronio Silva [1] to predict the Spanish electricity demand collected by the Spanish TSO, whose computational implementation uses a Mamdani fuzzy inference system, which directly processes a time series containing the demand data and its register. It should be noted that the database also includes forecasts made by the same Spanish TSO, so comparisons are made with these forecasts.

Keywords: FTS methods · PWFTS model · Electrical demand

1 Fuzzy Time Series Models

Fuzzy time series (FTS) can refer to a time series composed of fuzzy linguistic terms, or to the family of nonparametric forecasting methods introduced by Song and Chissom [2] based on the fuzzy set theory of Zadeh [3]. These methods allow handling numerical and non-numerical data, and have been used to predict the university enrollments [4], stock markets [5], tourism [6], electric load [7][8], among other phenomena.

The scalable models developed by Silva are 2 consensus methods that incorporate the main developments of the FTS literature for methods based on high-order rules and weighted rules. He also proposes a new model that uses probabilistic predictions considering 2 sources of uncertainty: fuzzy measures and stochastic behavior, being this a single-variable prediction model, which is also used as part of a multivariable prediction system.

Single-variable models: Since fuzzy logic analyzes the degrees of membership of variables to fuzzy sets, a great deal of knowledge can be extracted by studying the relationships that time series values have with the association to different fuzzy sets. The simplest models that perform this work by analyzing only the values belonging to the time series are the single-variable fuzzy models.

- **HOFTS** (*High Order Fuzzy Time Series*): This model, whose main parameter for the creation of rules is the order, does not weight the partition or the fuzzy sets, assuming that each of them has the same importance.
- **WHOFTS** (*Weighted High Order Fuzzy Time Series*): System that generates rules from the order and importance of fuzzy sets.
- **PWFTS** (*Probabilistc Weighted Fuzzy Time Series*): Model that in addition to considering the analysis of the previous models for training, uses the rules generated to change the importance of the fuzzy sets in a probabilistic way, which is applied again to the rules, feeding them back with new inferences.

Multivariate Models: Fuzzy models can also study the relationships between the values of a time series with related variables. However, these extra variables must have their own partition, since they correspond to values associated with a different phenomenom.

- **MVFTS** (*MultiVariable Fuzzy Time Series*): System that analyzes the relationships between variables using the linguistic variables of each fuzzy time series assuming that each fuzzy set has the same importance for each variable. It is similar to the HOFTS models since it only uses the linguistic terms to perform its analysis, but with the ability to have a more robust analysis when obtaining a prediction since it also uses information related to the target variable.
- **WMVFTS** (*Weighted Multivariable Fuzzy Time Series*: Extension of the previous system, where it adds importance to fuzzy sets.
- **FIG-FTS** (*Fuzzy Information Granular Fuzzy Time Series*): Model that has an emphasis on the relationships between variables through the use of a k-nearest neighbors system, in addition to incorporating weights between fuzzy sets and rules generated using the PWFTS system.

2 Dataset description

The information used for testing the different prediction models presented in this article is data collected by ENTSO-E (European Network of Transmission System Operators for Electricity) [10], a body of electricity managers that contains records on various variables related to the electricity market of several European countries. The main information in this database corresponds to the hourly Spanish electricity demand between 2015 and 2019 together with its date, the time of registration in format `yyyy/mm/dd - hh:mm:ss` and the forecast of electrical demand given by the organization itself. This dataset has 35064 samples, and for this work the sample characteristics used were those mentioned above, and was processed as a time series before being taken to the scalable FTS models.

3 Metodology and Implementation

The fuzzy logic systems used in this work are the scalable FTS models those described in section 1, developed by Petrônio Cândido de Lima e Silva in the `python pyFTS` [11] library, using Google Collaboratory environment to make the programation fo the systems. The programming was performed in Google Colaboratory with Nvidia 530.30.02 drivers and Python 3.9.16. . Given this, the computational process when using a fuzzy model use a Mamdani fuzzy inference system and therefore used the following procedure for training.

1. **Data preprocessing:** This is a generic step when using prediction models. Here all the cleaning and filtering of the information is done to take it to the models.
2. **Partition configuration:** This is the most important step when using fuzzy models, a precise analysis is performed on the target variable to know which is the most appropriate type of partition and number of fuzzy sets. The type of model to be used and its order is also configured.
3. **Fuzzyfication of the data:** The data are transferred to the fuzzy domain. At this point the fuzzy time series are generated.
4. **Generation of fuzzy rules:** Here the temporal transition rules are obtained, where they depend on the parameters with which the partition was configured. Depending on the model, the importance of the fuzzy sets (weighting models) and the relationships between the processed variables (multivariable models) are obtained.
5. **Forecast:** The predictions are made by taking a series of values whose length must be at least the order configured in the partition, in order to be able to make the prediction for the following periods.
6. **Defuzzyfication of the data:** Process opposite to fuzzyfication, at this point the data is returned to its original domain.

The data were cleaned of empty samples and since we were working with hourly electricity demand, **5 levels** were configured: Very Low, Low, Medium, High and Very High, where these sets were in turn divided into 7 sub levels, and the partition used was the grid with triangular membership function, giving the same range of membership to each fuzzy set depending on the value of the time series.

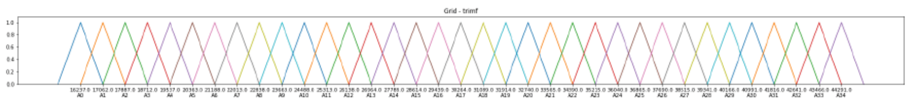

Fig. 1: Partition used for the data, 35 fuzzy sets were created with a minimum membership value of 15412 MW and a maximum of 45116, each interval had a length of 1650 MW.

The model was set up to apply this partition to the demand time series data, without using time windows, so that the series were processed all at once.

When multivariable models were used, the time variable was incorporated, and these series had their own partition. A lattice partition was also applied, but the records were processed by month and time, generating 2 fuzzy series with 12 and 24 fuzzy sets respectively, which were configured separately from the model. It should be noted that the partition applied to the demand data was the same as the one used in the single-variable models, except that it was not configured in the model, but together with the time partitions.

4 Results.

The predictions obtained by the single-variable fuzzy models are shown in Figure 2, where the shape of the forecasts obtained was similar over the entire time series of electricity demand. On the other hand, the multivariate fuzzy models had the behavior shown in Figure 3, where, as in the previous case, similar predictions were also obtained in other demand intervals.

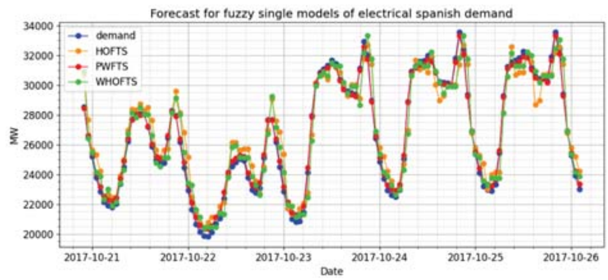

Fig. 2: Spanish electricity demand forecasts by fuzzy single variable models

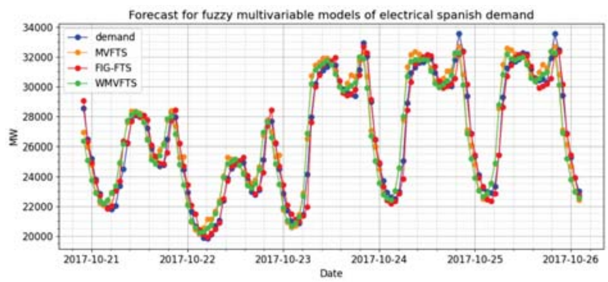

Fig. 3: Spanish electricity demand forecasts by fuzzy multivariable systems

The MAPE (Mean Absolute Percentage Error) was used as an error metric to evaluate the performance of the different systems.

Table 1: Performance of fuzzy multivariable prediction models

Model	Type	Order	MAPE
PWFTS	Fuzzy single-variable	1	0.8738
TSO Forecast	Not Applicable	Not Applicable	0.9590
FIG-FTS (k=2)	Fuzzy Multivariable	2	1.5421
FIG-FTS (k=3)	Fuzzy Multivariable	2	1.5866
WHOFTS	Fuzzy single-variable	3	2.3359
WHOFTS	Fuzzy single-variable	2	2.5408
HOFTS	Fuzzy single-variable	3	2.9445
WMVFTS	Fuzzy Multivariable	2	3.4820
MVFTS	Fuzzy Multivariable	1	3.6086
HOFTS	Fuzzy single-variable	2	3.6125

In general, the models were able to perform well in the one-step predictions, where the fuzzy single-variable systems did the job better and worse. This is partly due to the fact that although all of them have within their computational implementation the fuzzy logic theory, there are notorious differences in the generation of rules. Therefore, the values of the linguistic variable may differ quite a lot, which translates into different performance.

The fuzzy models that did not weight the sets were the ones that had a lower performance. This shows the great importance of this feature when setting up fuzzy models in general, since we will have more relevant fuzzy sets when generating the fuzzy rules and making the inferences, and later to make the demand predictions.

It should be noted that the single-variable models performed better if they were configured with a higher order, since the generation of their rules is strengthened if they have a larger amount of data to perform the analysis. However this does not happen with the PWFTS model, because this system generates its rules probabilistically, using the previous rules already generated to improve its analysis, so if a larger amount of samples are taken for the generation of a current rule then it reduces the amount of rules generated during its training, which leads to a reduction in its performance.

Unfortunately, it was not possible to find a description for the prediction method used by the Spanish TSO

5 Conclusion

The scalable models were able to make good one-step predictions from the demand data collected by the Spanish TSO. Although there were also configurations of these models that presented appreciable errors when forecasting demand such as the MVFTS and HOFTS models. However, these performances can be improved by changing the values of their hyperparameters, such as order.

The PWFTS model outperformed the predictions made by the Spanish TSO, thus showing the effectiveness of this system for time series predictions; however, this model is still under development and revision, so the results of the predictions obtained if this system is used with other databases should be carefully analyzed.

Although the use of this technique is not very well known, the work developed showed that it is very useful to work as prediction models. However, in order to demonstrate that these models can be the best for energy demand predictions, it is necessary to test them also with other databases.

References

1. P. Silva, *Scalable Models for probabilistic forecastng with fuzzy time series* Belo Horizonte - Minas Gerais Federal University of Minas Gerais (2019)
2. Q. Song and B. S. Chissom. Fuzzy time series and its models. *Fuzzy Sets and Systems*, 54 (3), pp:269-277, 1993. doi: 10.1016/0165-0114(93)90372-O.
3. L. A. Zadeh. Fuzzy Sets. *Information and Control* 8:338–353, 1965. DOI: 10.1016/S0019-9958(65)90241-X.
4. Q. Song and B. S. Chissorn. Forecasting enrollments with fuzzy time series-part II. *Fuzzy Sets and Systems* 62:1–8, 1994. doi: 10.1016/0165-0114(94)90067-1
5. M.-Y. Chen. A high-order fuzzy time series forecasting model for internet stock trading. *Future Generation Computer Systems*, 37:461–467, 2014. doi: 10.1016/j.future.2013.09.025.
6. M. H. Lee and H. Javedani. A Weighted Fuzzy Integrated Time Series for Forecasting Tourist Arrivals. In *International Conference on Informatics Engineering and Information Science* pp:206-217, Berlin Heidelberg, 2011. Springer. doi: 10.1007/978-3-642-25453-6_19
7. Z. Ismail, R. Efendi, and M. M. Deris. Application of Fuzzy Time Series Approach in Electric Load Forecasting. *New Mathematics and Natural Computation* 11(3), pp:229-248, 2015. doi:10.1142/S1793005715500076.
8. H. J. Sadaei, F. G. Guimarães, C. J. d. Silva, M. H. Lee, and T. Eslami. Short-term load forecasting method based on fuzzy time series, seasonality and long memory process. *International Journal of Approximate Reasoning*, 83:196–217, 2017. doi: 10.1016/j.ijar.2017.01.006.
9. S. Chen, *Forecasting enrollments based on fuzzy time series*, vol. 81, no. 3, pp. 311-319. 1996
10. N. Jhana, *Hourly energy demand generation and weather*, https://bit.ly/3nfphuw. Updated 2020
11. M. Intelligence and D. S. Laboratory, *pyFTS Quick Start*. 2018

Software Engineering and Databases