Fuzzy Interval Modelling based on Joint Supervision

Diego Muñoz-Carpintero
Institute of Engineering Sciences
Universidad de O'Higgins
Rancagua, Chile
diego.munoz@uoh.cl

Sebastián Parra
Department of Electrical Engineering
University of Chile
Santiago, Chile
sebastian.parra.f@ug.uchile.cl

Oscar Cartagena
Department of Electrical Engineering
University of Chile
Santiago, Chile
oscar.cartagena@ing.uchile.cl

Doris Sáez
Department of Electrical Engineering
University of Chile
Santiago, Chile
dsaez@ing.uchile.cl

Luis G. Marín
Department of Electrical Engineering
Universidad de Los Andes
Bogotá, Colombia
l.marinc@uniandes.edu.co

Igor Škrjanc
Faculty of Electrical Engineering
University of Ljubljana
Ljubljana, Slovenia
Igor.Skrjanc@fe.uni-lj.si

***Abstract*—This paper presents a new methodology for Prediction Interval (PI) construction based on a modified Takagi-Sugeno fuzzy system trained with a joint Supervision loss function. Given a desired coverage level, this model is capable of providing predictions of the expected value of the system along with the interval bounds. This methodology is tested by simulation experiments using a dataset containing real temperature data from a rural community in southern Chile. The proposed model was compared with a state-of-the-art Takagi-Sugeno Fuzzy Numbers model. It was shown that the Joint Supervision method manages to obtain slightly superior results to the Fuzzy Numbers approach while greatly reducing the complexity of the training loss function. Additionally, since the proposed model was trained using Particle Swarm Optimization, further performance improvements could be made by employing gradient-based optimization algorithms, since they are compatible with the Joint Supervision loss function.**

***Index Terms*—prediction interval, fuzzy systems, joint supervision**

I. Introduction

Predictive models have been a crucial tool used to estimate partly stochastic phenomena found in scientific and engineering applications. However, the emergence of decision-making problems such as microgrid operation, where the main challenge is to decide whether to rely on renewable energy sources or to utilize a diesel-based alternative based on predictions of renewable energy availability (for example, wind speed [1] or solar energy [2]) and electric load [3], has introduced the necessity of quantifying prediction uncertainty in addition to the desire of obtaining precise estimations, which makes traditional point regression models insufficient. In this context, Confidence Intervals (CIs) and Prediction Intervals (PIs) have been extensively researched as an alternative for uncertainty characterization by mean of interval models rather than only crisp prediction. Confidence intervals are used to capture uncertainties in the unknown parameters of a model and prediction intervals are used to capture uncertainties in random variables yet to be observed and provide a probability that the random variable will be within a given interval [4]–[7]. The predicted outputs are intervals that represent the most likely region defined by the upper and lower bounds of the interval to which the output (targets) of the uncertain phenomena will belong. Prediction intervals aim to maximize the amount of data inside the bounds, as well as to be as sharpest interval possible [8]–[10].

There have been many reported approaches to construct prediction intervals in literature [11]–[15], but Computational Intelligence models, such as fuzzy systems [17] and neural networks [18], have become increasingly popular [8], [9], [16] due to their capacity to capture nonlinear behaviour from complex systems and the availability of sufficient input-output training data. Other strategies for interval models include [11]–[13], [15]. The work of [11] uses regression tree methods for the construction of interval models. An ensemble machine learning approach was proposed in [12] to enhance both the reliability and width of prediction intervals. In the work of [13] the prediction interval for wind power was developed using the kernel extreme learning machine (KELM) method. In [15], an improved bootstrap method was proposed for constructing prediction intervals using extreme gradient boosting (XGB) as the base model.

Among computational intelligence alternatives, Takagi-Sugeno fuzzy interval models stand out due to their high interpretability, as these estimators are built under the intuition that nonlinear systems can be approximated as a combination of various local linear models. Using these models, many

This study was partially supported by Instituto de Sistema Complejos ISCI (ANID PIA/BASAL AFB180003), Solar Energy Research Center SERC-Chile (ANID/FONDAP/15110019) and FONDECYT 1170683 "Robust Distributed Predictive Control Strategies for the Coordination of Hybrid AC and DC Microgrids", and CONICYT PAI Convocatoria Nacional Subvención a Instalación en la Academia Convocatoria 2009 PAI77190021.

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

techniques have been proposed for PI construction systems, where the uncertainty is modelled in the antecedent, in the consequent or in both parts of the fuzzy rules [19], [20]. The main challenges reside in being able to create intervals as slim as possible, given a desired PI coverage level. Prediction intervals with narrower width give more accurate information about the uncertainty phenomena. However, a width that is too narrow might compromise the amount of measured data that belong to the interval [8].

In the specialized literature, there are several methods reported to construct fuzzy interval models, for instance, the covariance method, which uses the error between the prediction of the model and the real data to generate the interval [10]. Several applications based on this approach have been reported, such as: a model for the pH-titration curve [21], a one-day-ahead prediction models for renewable generation sources and electric demand [16], an approach to indoor localization [22], and a model for a waste-water treatment plant, which exhibit very nonlinear behaviour [23]. Previous studies are based on covariance method to develop the fuzzy interval models, however, in these studies, the normally distributed noise with zero mean is an a priori assumption.

Other fuzzy models have been proposed to generate prediction intervals without any assumptions on the distribution of the data because these kinds of models can naturally provide the interval with a type-1 fuzzy set at the consequent. For instance, in [24] the dispersion of the output (uncertainty) was considered in the design of the fuzzy model to obtain the prediction interval in an active way. The parameters of premises and consequences of the fuzzy model were found using an improved teaching-learning-based optimization algorithm (ITLBO) that minimize a multi-objective cost function. In [25] a type-2 fuzzy system was proposed to develop interval models. The left and right points of the type-reduced set were defined as the lower and upper bounds of the prediction interval. However, the parameters were obtained using a metric as optimality criteria of the prediction interval that fails when an extremely narrow interval width is obtained and therefore, in these cases, the coverage probability can be very low, as was discussed in [26], and [27].

On the other hand, in the studies of [28], [29] and [30] have used an optimization procedure that minimizes the maximal estimation error between the data and the proposed model output to find the parameters of fuzzy interval model. As a result of the optimization problem, the lower and upper bounds of the interval are obtained, however, the constraints do not include metrics to guarantee the desired value of coverage probability with minimum interval width.

In this work, multi-output fuzzy Takagi-Sugeno models are proposed to develop Prediction Interval models that are both high-quality and interpretable while also being suited for online applications. With this approach, the crisp prediction and the interval are generated under a single fuzzy model.

The main contribution of this work is the development of a methodology to generate fuzzy prediction interval models using a Joint Supervision loss function, initially presented in [10] as a function to be used in neural predictive models to obtain fast and precise PIs, and later used in [14] for obtaining predictions interval models with Long Short Term Memory neural networks. The proposed approach tries to obtain Prediction Interval models that generate the most information in terms of the relationship between the width of the interval and the quantity of measured data that fall within the interval defined by the model.

In order to test this approach, experiments were run on an experimental dynamical system using a dataset containing real temperature data measured on the rural José Painecura mapuche community located in Región de la Araucanía, Chile. Using this data, the performance of the proposed model was studied by comparing Prediction Interval average width and RMS error with a Takagi-Sugeno Fuzzy Numbers PI model [8], which has shown state-of-the-art results on PI quality, trained on the same dataset. The results show that the prediction intervals generated by the proposed approach in this study are more accurate than those generated by the method used for comparison. The above, because the proposed method obtains a narrower interval width and this interval contains the desired percentage of measured data inside the bounds. This is important since an accurate prediction of the uncertainty provides both more information about the phenomena modelled and more useful information from a decision-making point of view.

The paper is organized as follows: the fuzzy Takagi-Sugeno model architecture and the original proposal for neural Joint Supervision are explained in Section II. In Section III, the Takagi-Sugeno Joint Supervision proposed formulation is explained, describing both its model architecture and training procedure. Then, in Section IV the experimental test is presented, where the main parameters and configurations considered during model training are discussed and results are shown and analyzed. Finally, the main conclusions and contributions of this work are discussed in Section V.

II. Background: Fuzzy Modelling and Neural Joint Supervision

A. *Fuzzy Takagi-Sugeno Predictive Modeling*

Takagi-Sugeno fuzzy models are often referred as universal approximators of non-linear functions, since they can uniformly approximate any continuous function with arbitrarily high precision by using a sufficiently large number of local sub-models. There are three main building blocks common to all fuzzy models. First, the crisp input introduced to the model undergoes a fuzzification procedure that converts it into fuzzy values. Then, with the help of a fuzzy ruleset, an inference system calculates a degree of activation for each rule, which tries to determine which local sub-models best fit the input data, so that a fuzzy output can be obtained. Finally, a defuzzification procedure is applied to convert the fuzzy output into crisp values.

For Takagi-Sugeno models, it is assumed that all local sub-models are linear models, so fuzzy rules are expressed as:

$$R^j : \text{If } z_1(k) \text{ is } A_1^j \text{ and ... and } z_n(k) \text{ is } A_n^j \text{ then} \quad (1)$$
$$\hat{y}^j(k) = \theta_0^j + \theta_1^j z_1(k) + ... + \theta_n^j z_n(k),$$

for $j = 1, ..., M$, where A_i^j are the fuzzy sets, $\hat{y}^j(k)$ is the local output and $\theta^j = [\theta_0^j, \theta_1^j, ..., \theta_n^j]$ is the consequence parameter, all associated with rule j, M is the total number of rules and $z(k) = [z_1(k), z_2(k), ..., z_n(k)]$ is the input vector (normally, for predictive models these values correspond to past observations). Using this structure, Takagi-Sugeno models can calculate a degree of activation $h_j(k)$ for each rule j, along with its normalized variant $\beta_j(k)$, according to

$$h_j(k) = \prod_{i=1}^{n} \mu_{A_i^j}(z_i(k)), \quad (2)$$

$$\beta_j(k) = \frac{h_j(k)}{\sum_{j=1}^{M} h_j(k)}, \quad (3)$$

where $\mu_{A_i^j}$ is the fuzzy activation function for fuzzy set A_i^j. Finally, using this ruleset and inference mechanism, a crisp output can be calculated as

$$\hat{y}(k) = \sum_{j=1}^{M} \beta_j(k) \hat{y}^j(k), \quad (4)$$

where $\hat{y}(k)$ is the total predicted output at sample k.

For this work, this will be the base model structure used to develop the proposed methodology.

B. *Neural Joint Supervision*

Originally presented in [10], the Joint Supervision method provides a methodology that can be applied to standard neural predictive models to extend them into interval models. To accomplish this, first the neural architecture is modified to incorporate two additional outputs: one for a prediction interval upper bound, and one for a lower interval bound, so that model outputs can be expressed as $\hat{y}_{upper}(k)$, $\hat{y}_{crisp}(k)$, and $\hat{y}_{lower}(k)$. Note that this procedure differs from training three separate networks in that intermediate layers (and in consequence, their corresponding parameters) will be shared among all outputs.

The model is trained using novel cost functions, associated with each output, based on the idea that if a training sample falls outside the predicted interval, then it should be penalized proportionally to its MSE. This is achieved by splitting the training loss function in two components: a **Classical Loss Function** (L_s), which is the traditional MSE loss given by

$$L_s^l = \frac{1}{N} \sum_{k=1}^{N} e_l^2(k), \quad (5)$$

$$e_l(k) = y(k) - \hat{y}_l(k), \quad (6)$$

and is the same for all outputs ($l = \{upper, crisp, lower\}$), and an **Interval Loss Function** (L_I), which is given by

$$L_I^{upper} = \frac{1}{N} \sum_{k=1}^{N} \frac{\text{sgn}(e_{upper}(k)) + 1}{2} e_{upper}(k)^2, \quad (7)$$

$$L_I^{crisp} = 0, \quad (8)$$

$$L_I^{lower} = \frac{1}{N} \sum_{k=1}^{N} \frac{\text{sgn}(-e_{lower}(k)) + 1}{2} e_{lower}(k)^2, \quad (9)$$

where

$$sgn(x) := \begin{cases} -1 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0, \end{cases} \quad (10)$$

and only affects the interval outputs (i.e. $l = \{upper, lower\}$). The main idea behind these functions is that L_I^{upper} will penalize all data that fall above the upper interval bound with an amount proportional to the squared distance between both values, while L_I^{lower} will apply the corresponding penalization for data that fall below the predicted lower interval bound. Finally, the total loss function for each model output is built by adding the two corresponding components, with the inclusion of an additional hyperparameter λ common to all three outputs, which is used to control the importance of the interval loss function during the training procedure, allowing for the user to be able to control the final interval size:

$$L_{total}^l = L_s^l + \lambda L_I^l. \quad (11)$$

By adding these two components, the Joint Supervision function manages to mathematically represent the two contradicting objectives found in PI construction, where the interval loss function attempts to have most point within the PI (increasing PI width), while the classical loss function seeks to converge all outputs to the expected data value (reducing PI width). This way, it is the tuning of the λ hyperparameter that will provide a suitable compromise that will determine average interval width and coverage level.

In fact, the λ hyperparameter tuning is performed through an iterative process, where an initial value (usually $\lambda_0 = 0.001$) and a fixed step δ are set. Then, a model is trained for λ_0 and the PI coverage level is measured. If coverage is less than desired (based on user preference, although usual coverage level values range from 90% to 95%), then a new hyperparameter value is set as $\lambda_i = \lambda_{i-1} + \delta$, where λ_{i-1} is the hyperparameter value used on the previous iteration, and the model is retrained, repeating this process until the desired coverage level is reached. If, on the other hand, coverage level is greater or equal than the desired value, the model will be retrained several times using the same configuration, measuring the average interval width on each iteration and sticking with the model that showcases the lowest width. This final step is done in order to reduce uncertainty due to parameter initialization.

One of the main advantages of the Joint Supervision method when compared to other neural PI models is that, while managing to build accurate PIs, the Joint Supervision cost function's simplicity makes it compatible with traditional neural network training algorithms based on backpropagation and stochastic gradient descent, which eases implementation and significantly boosts training speed.

III. Fuzzy Takagi-Sugeno Joint Supervision Interval Models

This work proposes a methodology to adapt the structure of Takagi-Sugeno fuzzy models in order to be compatible with the Joint Supervision method for PI construction, which requires a multi-output architecture that can take advantage of some degree of parameter sharing.

First, a new multi-output formulation for Takagi-Sugeno models was developed with the intention to preserve the weight sharing principle found in multi-output neural networks. Specifically, a three output fuzzy model is considered where local and total outputs can be represented as

$$\hat{y}_l^j(k) = [1 \ z(k)^T]\theta_l^j, \tag{12}$$

$$\hat{y}_l(k) = \sum_{j=1}^{M} \beta_j(k)\hat{y}_l^j(k), \tag{13}$$

$l = upper, crisp, lower$ identifies the three different outputs, $j = 1, \ldots, M$ identifies the j-th rule and θ_l^j is the consequence parameter vector for the j-th rule of model l. As with multi output neural architectures, this approach differs from training three separate model instances in that, in this formulation, all model outputs share the same fuzzy ruleset and, as such, will also share the same degrees of activation β_j. This formulation not only makes mathematical sense, since Takagi-Sugeno models can be seen as a network model where the final and only separated layer corresponds to the product between the input vector and the corresponding consequence parameters, but it also makes sense from a practical perspective, since the fuzzy sets defined on each rule define an operation zone on the input data where the defined local outputs are valid and sharing this parameters follows the intuition that all local outputs associated to a rule should be related to a common operation zone on the input data. In order to better illustrate this intuition, figure 1 provides a comparative diagram of a neural Joint Supervision model structure and the proposed Takagi-Sugeno Joint Supervision architecture.

The training procedure is described next, and is illustrated in Figure 2. First, a sensitivity analysis is run on the training data in order to determine the optimal set of regressors (which compose the model input vector). Then, a fuzzy clustering algorithm is run on the same data to determine the optimal number of fuzzy rules along with the membership functions of the fuzzy sets. After this, the consequence parameter vectors θ_l^j for each output $l = upper, crisp, lower$ are obtained by optimizing the Joint Supervision cost functions as presented in (11):

$$J_{upper} = L_s^{upper} + \lambda L_I^{upper} \tag{14}$$

$$J_{crisp} = L_s^{crisp} \tag{15}$$

$$J_{lower} = L_s^{lower} + \lambda L_I^{lower} \tag{16}$$

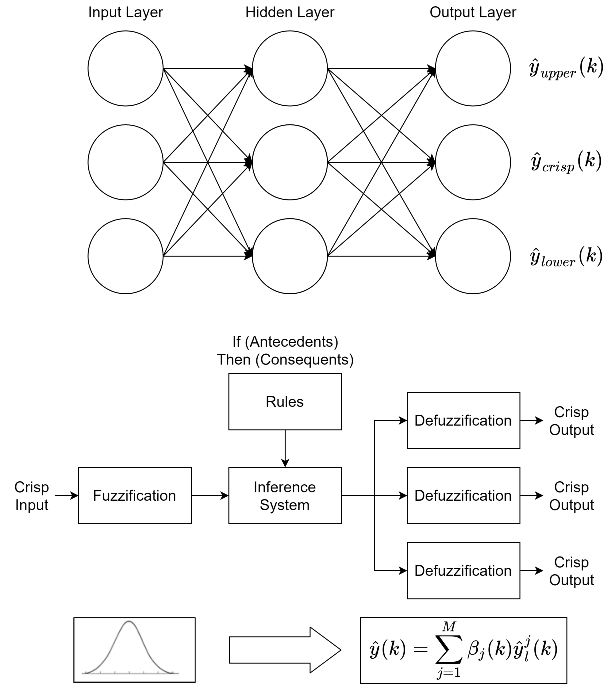

Fig. 1. Model architecture of a neural Joint Supervision interval model (top). Model architecture of proposed Takagi-Sugeno Joint Supervision interval model (bottom).

where $y(k)$ corresponds to the k-th sample of the training set. This defines θ_{crisp}^j, and consequently the crisp model. However, the upper and lower models will be defined along with the tuning of λ, for which the optimization above is a starting point performed with $\lambda = \lambda_0$.

Finally, hyperparameter tuning is executed as in neural Joint Supervision, going through an iterative process trying different values of λ, with additional retraining step to combat parameter initialization uncertainty, until the the desired coverage level is reached. It is important to note that in this work the value of λ was adjusted through a logarithmic search. In this implementation, on each iteration the algorithm first trains a model, with its corresponding λ_i value, using Particle Swarm Optimization. Once the model is obtained, the algorithm will check the if the measured PI coverage level is in the vicinity of the desired coverage value. If the condition is satisfied, then the current configuration is saved and the algorithm proceeds to the second retraining phase (the one that reduces uncertainty due to parameter initialization), which is executed identically to traditional Joint Supervision. If the condition is not satisfied, then the model is retrained using a new λ_i value given by the expression

$$\lambda_i = \begin{cases} 2\lambda_{i-1} & \text{if } \lambda_{up} = 0 \wedge CL_{measured} < CL_{desired} \\ \frac{\lambda_{i-1}+\lambda_{low}}{2} & \text{if } CL_{measured} > CL_{desired} \\ \frac{\lambda_{i-1}+\lambda_{up}}{2} & \text{if } CL_{measured} < CL_{desired}, \end{cases} \tag{17}$$

where CL stands for coverage level, and λ_{up} and λ_{low} represent pivot values that are initialized as zero and correspond

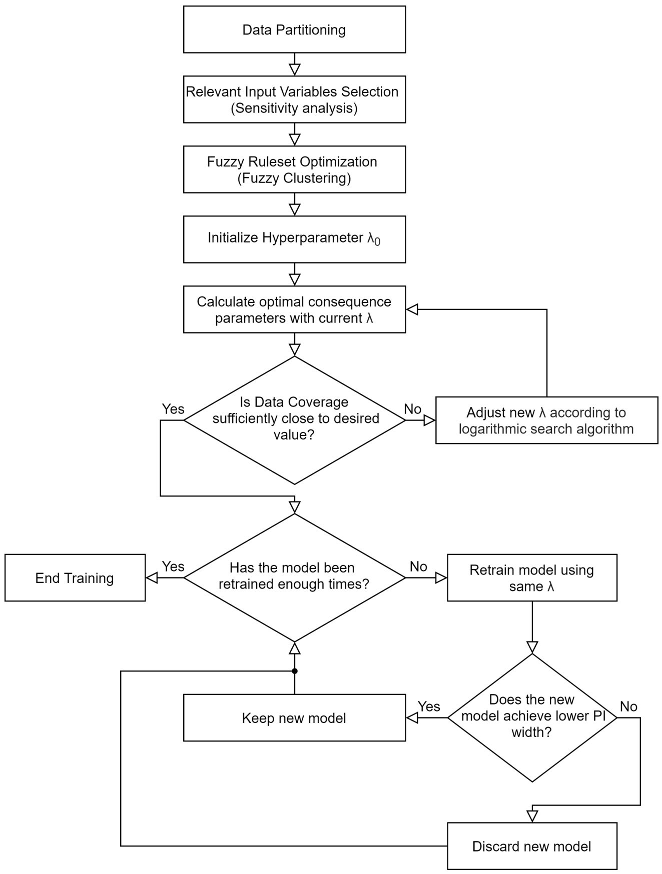

Fig. 2. Proposed Takagi-Sugeno Joint Supervision training routine.

to the lowest/highest λ value that overshoots/undershoots the desired coverage, respectively. Coverage level is quantified by the Prediction Interval Coverage Probability (PICP) [31] [32] index, given by

$$PICP = \frac{1}{N}\sum_{k=1}^{N}\delta_k, \tag{18}$$

where $\delta_k = 1$ if $\hat{y}_{lower}(k) \leq y(k) \leq \hat{y}_{upper}(k)$ and $y(k)$ represents the k-th training sample.

IV. Experimental Tests

The proposed model was studied with simulation experiments using a dataset consisting of real temperature values measured in the José Painecura mapuche community located in Región de la Araucanía, Chile. For comparative purposes, PI quality measurements were compared with a Fuzzy Numbers Takagi-Sugeno interval model [8], which has reported state-of-the-art results on PI construction.

In order to properly train the models, data was split into three sets:

1) **Traning set**, composed of 60% of the data, used to train the fuzzy clustering algorithm along with the PSO algorithm for parameter optimization
2) **Testing set**, composed of 20% of the data, used to execute the sensitivity analysis for model structure optimization
3) **Validation set**, composed of 20% of the data, used to measure and compare model performance

Once data was partitioned, models proceeded to be trained with the caution of keeping the same set of regressors for both models, so that differences in model performance could only be attributed to model structure. The baseline Takagi-Sugeno architecture consisted of 3 fuzzy rules with gaussian membership functions as antecedents and first order linear models as consequents. The number of fuzzy rules was determined using the Gustafson-Kessel fuzzy clustering algorithm on the training data. Once training finished, 90% confidence prediction intervals were calculated for 1-step to 6-steps ahead predictions, where the Fuzzy Numbers model was tuned to obtain a maximum error margin of $\pm 0.006\%$ coverage level on the training set, while the Joint Supervision model used a margin of $\pm 0.01\%$ for hyperparameter tuning. For this experiment, the optimal model structure was found to consist of 5 regressors and 3 fuzzy rules.

Models were evaluated based on their average point prediction precision (RMSE) along with average PI quality, which was measured by computing Prediction Interval Coverage Probability (PICP) as shown in (18), and Prediction Interval Mean Average Width (PINAW) [31] [32], given by

$$\text{PINAW} = \frac{1}{NR}\sum_{k=1}^{N}\hat{y}_{upper}(k) - \hat{y}_{lower}(k) \tag{19}$$

where $R = \max(y_{upper}(k)) - min(y_{lower}(k))$, which quantify the average coverage level and width of the intervals, respectively.

Tables I and II show the measured performance values for the 1-step to 6-step ahead predictions on the Joint Supervision and Fuzzy Numbers models, respectively. Using these criteria, a good PI can be recognized if it manages to achieve PICP values that are as close as possible (usually $\pm 1\%$) to the desired coverage level (in this case, 90%), while also displaying as low

TABLE I
JOINT SUPERVISION METHOD PERFORMANCE FOR 1-STEP TO 6-STEPS AHEAD PREDICTIONS

	1-Step	**2-Steps**	**3-Steps**	**4-Steps**	**5-Steps**	**6-Steps**
PICP	89.59%	89.03%	89.11%	87.61%	89.61%	88.93%
PINAW	5.19%	8.00%	10.08%	11.36%	14.34%	15.38%
RMSE	0.29	0.48	0.6	0.69	0.78	0.87

TABLE II
FUZZY NUMBERS METHOD PERFORMANCE FOR 1-STEP TO 6-STEPS AHEAD PREDICTIONS

	1-Step	**2-Steps**	**3-Steps**	**4-Steps**	**5-Steps**	**6-Steps**
PICP	92.69%	93.54%	93.58%	93.41%	92.08%	92.16%
PINAW	5.85%	10.56%	13.58%	15.91%	17.18%	20.41%
RMSE	0.29	0.48	0.6	0.69	0.78	0.87

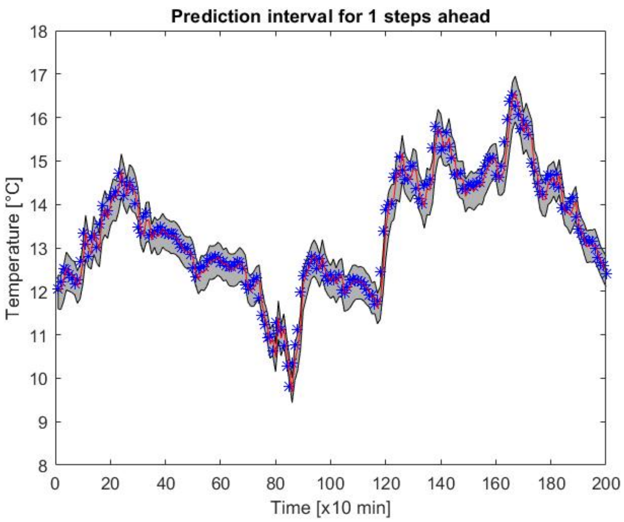

Fig. 3. Joint Supervision Prediction Interval for 1-step predictions on the validation set.

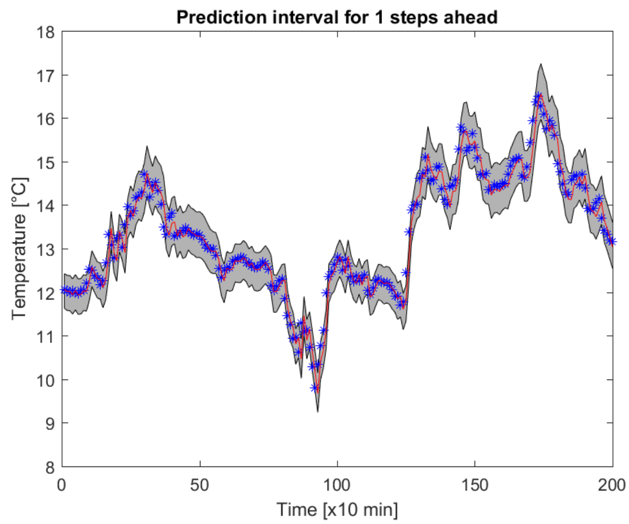

Fig. 4. Fuzzy Numbers Prediction Interval for 1-step predictions on the validation set.

as possible PINAW values. It is important to note that, since the coverage and width objectives are conflicting, exceeding the PICP condition means that a lower PINAW value would be theoretically achievable if the PI model were to be adequately tuned, while the opposite is true on cases where the obtained coverage falls below the desired value.

By observing Tables I and II, it can be seen that, in the first place, the Joint Supervision model manages to stay within the $\pm 1\%$ PICP range for a majority of the predicted models, while the Fuzzy Numbers approach shows a tendency to exceed the desired value by a significant margin, which can have a negative impact on PINAW performance. In fact, by observing the registered PINAW values along with Figures 3-8, it can be seen that the Joint Supervision model matches or outperforms the Fuzzy Numbers approach on this aspect. It is important to note that in both methodologies, the exact same structure and training procedure is used for the crisp model. Therefore the crisp models are the same, which results the RMSE performance being exactly the same on both models. This analysis shows that the proposed Joint Supervision model preliminarily achieves similar, possibly superior, performance to known state-of-the-art interval models.

Additionally, Figures 3, 5 and 7 show 1-step, 3-step and 6-step predictions for the Joint Supervision model, respectively, while Figures 4, 6 and 8 show respective plots for the Fuzzy Numbers model. By observing these predictions, it can be noted that the Joint Supervision model consistently obtains narrower intervals, which is consistent with the results observed in Tables I and II.

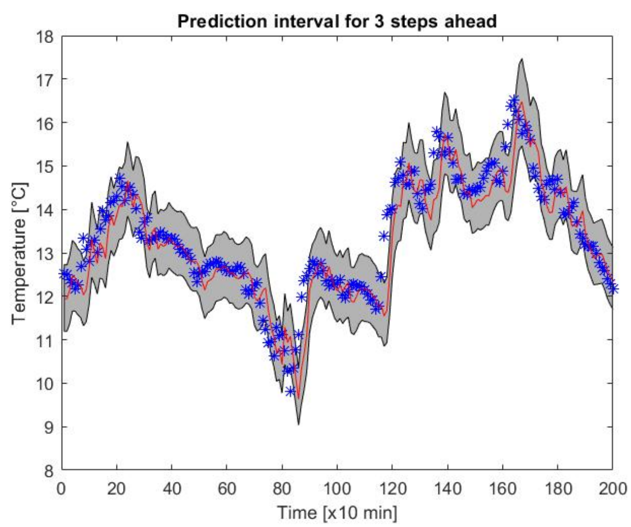

Fig. 5. Joint Supervision Prediction Interval for 3-step predictions on the validation set.

V. Conclusions

A methodology for PI construction based on modified multi-output Fuzzy Takagi-Sugeno models using a Joint Supervision cost function was proposed. The model was built in an attempt to take advantage of the simplicity, convexity an overall good interval quality of the Joint Supervision method's cost function along with the high interpretability of Takagi-Sugeno systems. The proposed approach was tested with real data and compared with a Fuzzy Numbers interval model which can build precise state-of-the-art PIs using a more complex cost function, implementing the optimization of both methods with PSO. It was found that both models obtain comparable results, although the Fuzzy Numbers approach showcased a tendency to exceed the desired the coverage level on the validation set, which, considering the model coverage was precisely tuned during the training phase, would hint to an imprecision in the interval model itself, while the Joint Supervision method obtained a smaller, more precise, margin for the coverage level values, therefore producing, on average, slightly narrower intervals than the Fuzzy Numbers approach. Because of these results,

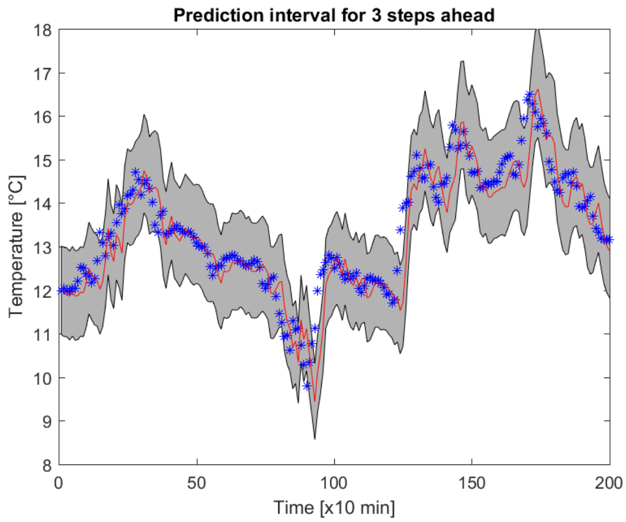

Fig. 6. Fuzzy Numbers Prediction Interval for 3-step predictions on the validation set.

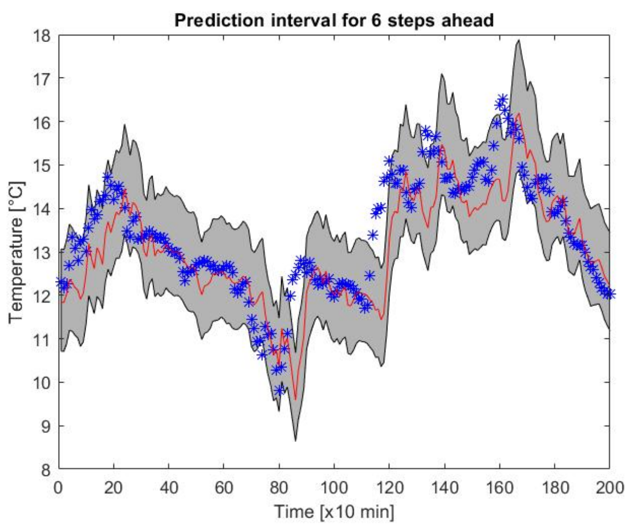

Fig. 7. Joint Supervision Prediction Interval for 6-step predictions on the validation set.

it can be concluded that the proposed fuzzy Takagi-Sugeno Joint Supervision methodology preliminarily has comparable performance to other state-of-the-art solutions, although additional experiments are required to confirm if the proposed model is capable of outperforming them. Additionally, there is space for improvement in the Joint Supervision method's computational cost: Since PSO was used for the optimization procedure in this work, improvements could be achieved by employing gradient-based methods that can take advantage of the convex Joint Supervision cost function, which should reduce parameter initialization uncertainty and overall training times. Furthermore, the need of repeated retraining during the hyperparameter tuning phase of the algorithm could be revised to explicitly introduce the desired PICP into the Joint Supervision cost function in order to eliminate the need for a

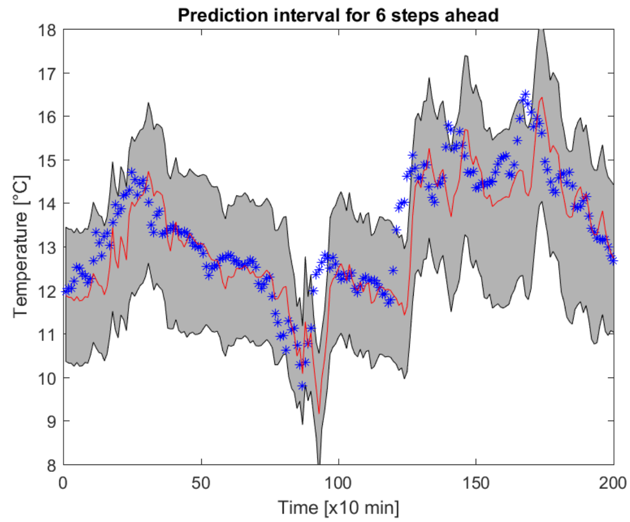

Fig. 8. Fuzzy Numbers Prediction Interval for 6-step predictions on the validation set.

costly iterative process during the training phase.

References

[1] P. Pinson and G. Kariniotakis, "Conditional Prediction Intervals of Wind Power Generation", IEEE Transactions on Power Systems, vol. 25, no. 4, pp. 1845-1856, Nov. 2010.

[2] I. Galván, J. Valls, A. Cervantes and R. Aler, "Multi-objective evolutionary optimization of prediction intervals for solar energy forecasting with neural networks", Information Sciences, vol. 418-419, pp. 363-382, 2017.

[3] A. Khosravi, S. Nahavandi and D. Creighton, "Load Forecasting and Neural Networks: A Prediction Interval-Based Perspective", Computational Intelligence in Power Engineering, pp. 131-150, 2010.

[4] R. Dybowski and S. J. Roberts. "Confidence intervals and prediction intervals for feed-forward neural networks", page 298–326. Cambridge University Press, 2001.

[5] T. Heskes. "Practical confidence and prediction intervals.Advances in Neural Information Processing Systems", 9:176–182, Oct. 1997.

[6] M. Ramezani, M. Bashiri, and A. C. Atkinson. "A goal programming-topsis approach to multiple response optimization using the concepts of non-dominated solutions and prediction intervals", Expert Systems with Applications, 38(8):9557–9563, 2011.

[7] A. C. Rencher and G. B. Schaalje."Linear models in statistics", John Wiley & Sons,Inc, USA, 2 edition, 2008.

[8] L.G. Marín, N. Cruz, D. Sáez, M. Sumner and A. Núñez, "Prediction interval methodology based on fuzzy numbers and its extension to fuzzy systems and neural networks", Expert Systems with Applications, vol. 119, pp. 128-141, 2019.

[9] H. M. D. Kabir, A. Khosravi, M. A. Hosen and S. Nahavandi, "Neural Network-Based Uncertainty Quantification: A Survey of Methodologies and Applications", in IEEE Access, vol. 6, pp. 36218-36234, 2018.

[10] N. Cruz, L. G. Marín and D. Sáez, "Neural Network Prediction Interval Based on Joint Supervision", 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, 2018, pp. 1-8.

[11] Cyril Voyant, Fabrice Motte, Gilles Notton, Alexis Fouilloy, Marie-Laure Nivet, Jean-Laurent Duchaud, Prediction intervals for global solar irradiation forecasting using regression trees methods, Renewable Energy, Volume 126, 2018, Pages 332-340.

[12] W. Zhang, H. Quan and D. Srinivasan, "Prediction Interval Construction for Electric Load and Wind Power via Machine Learning," 2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), Singapore, 2018, pp. 716-721.

[13] Hu, M.; Hu, Z.; Yue, J.; Zhang, M.; Hu, M. A Novel Multi-Objective Optimal Approach for Wind Power Interval Prediction. Energies 2017, 10, 419.

[14] N. Cruz, L. G. Marín and D. Sáez, "Prediction Intervals With LSTM Networks Trained By Joint Supervision," 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1-8.
[15] Kaiwen Li, Rui Wang, Hongtao Lei, Tao Zhang, Yajie Liu, Xiaokun Zheng, Interval prediction of solar power using an Improved Bootstrap method,Solar Energy,Volume 159,2018,Pages 97-112.
[16] D. Sáez, F. Ávila, D. Olivares, C. Cañizares and L. Marín, "Fuzzy Prediction Interval Models for Forecasting Renewable Resources and Loads in Microgrids", in IEEE Transactions on Smart Grid, vol. 6, no. 2, pp. 548-556, March 2015.
[17] R. Babuska, "Fuzzy modeling for control", Boston: Kluwer Academic Publishers, 1998.
[18] M. Nørgaard, "Neural networks for modelling and control of dynamic systems", London: Springer, 2010.
[19] J.M. Mendel, "Uncertain rule-based fuzzy logic systems: Introduction and new directions.Springer International Publishing", 2nd edition, 2017.
[20] L.G Marín, F. Valencia, F., and D. Sáez, "Prediction interval based on type-2 fuzzy systems for wind power generation and loads in microgrid control design", in Proceedings of the IEEE international conference on fuzzy systems (FUZZ-IEEE), Vancouver, BC, 2016, pp. 328–335.
[21] I. Škrjanc. Fuzzy confidence interval for pH titration curve. Applied Mathematical Modelling, 35(8):4083–4090, 2011.
[22] S. Tomažič, D. Dovžan and I. Škrjanc, "Confidence-Interval-Fuzzy-Model-Based Indoor Localization", in IEEE Transactions on Industrial Electronics, vol. 66, no. 3, pp. 2015-2024, March 2019.
[23] I. Škrjanc, "Confidence interval of fuzzy models: An example using a waste-water treatment plant", Chemometrics and Intelligent Laboratory Systems, Volume 96, Issue 2, Pages 182-187,2009.
[24] F. Veltman, F., L.G. Marín, D. Sáez, L. Gutierrez, and A. Nuñez, "Prediction interval modelling tuned by an improved teaching learning algorithm applied to load forecasting in microgrids", In Proceedings of the IEEE symposium series on computational intelligence (pp. 651–658). Cape Town, 2015.
[25] A. Khosravi, S. Nahavandi, "An interval type-2 fuzzy logic system-based method for prediction interval construction", Applied Soft Computing, Volume 24, Pages 222-231, 2014.
[26] C. Wan, Z. Xu, J. Østergaard, Z.Y. Dong, and K.P. Wong,"Discussion of combined nonparametric prediction intervals for wind power generation", IEEE Transactions on Sustainable Energy, 5(3),2014.
[27] A. Khosravi, and S. Nahavandi, Closure to the Discussion of "Prediction Intervals for Short-Term Wind Farm Generation Forecasts" and "Combined Nonparametric Prediction Intervals for Wind Power Generation" and the Discussion of "Combined Nonparametric Prediction Intervals for Wind Power Generation", IEEE Transactions on Sustainable Energy, 5(3), 1022–1023, 2014.
[28] I. Škrjanc, Sašo Blažič, Osvaldo Agamennoni, "Identification of dynamical systems with a robust interval fuzzy model", Automatica, Volume 41, Issue 2, Pages 327-332, 2005.
[29] I. Skrjanc, S. Blazic and O. Agamennoni, "Interval Fuzzy Model Identification Usingl_infty-Norm", in IEEE Transactions on Fuzzy Systems, vol. 13, no. 5, pp. 561-568, Oct. 2005.
[30] I. Skrjanc, S. Blazic and O. E. Agamennoni, "Interval fuzzy modeling applied to Wiener models with uncertainties", in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 35, no. 5, pp. 1092-1095, Oct. 2005.
[31] A. Khosravi, S. Nahavandi, D. Creighton and A. F. Atiya, "Lower Upper Bound Estimation Method for Construction of Neural Network-Based Prediction Intervals", in IEEE Transactions on Neural Networks, vol. 22, no. 3, pp. 337-346, March 2011.
[32] A. Khosravi, S. Nahavandi, D. Creighton and A. F. Atiya, "Comprehensive Review of Neural Network-Based Prediction Intervals and New Advances", in IEEE Transactions on Neural Networks, vol. 22, no. 9, pp. 1341-1356, Sept. 2011.