K Nearest Neighbour Optimal Selection in Fuzzy Inductive Reasoning for Smart Grid Applications

Sergio Jurado
Soft Computing research group at Intelligent Data Science and Artificial Intelligence Research Center Universitat Politècnica de Catalunya
Barcelona, Spain
s.juradogomez@gmail.com

Àngela Nebot
Soft Computing research group at Intelligent Data Science and Artificial Intelligence Research Center Universitat Politècnica de Catalunya
Barcelona, Spain
angela@cs.upc.edu

Francisco Mugica
Soft Computing research group at Intelligent Data Science and Artificial Intelligence Research Center Universitat Politècnica de Catalunya
Barcelona, Spain
fmugica@cs.upc.edu

Energy forecasting has been an area of great interest in the last years. It unlocks, not only the Smart Grid's potential with load balancing but also new business models and added value services. To achieve an accurate, robust and fast prediction, model's parametrization is key and becomes a bottleneck in the value-chain. In this article, we present an improved version of Flexible Fuzzy Inductive Reasoning (Flexible FIR) that selects the most optimal number of nearest neighbours during FIR prediction phase, called K nearest neighbour Optimal Selection (KOS). To this end, a real smart grid forecasting application, i.e. electricity load forecasting, has been chosen in this study. The results show that the best forecasting accuracy, on average, is when the KOS is used on Flexible FIR. While with KOS the optimal parameter k is found online, without it is not, which increases the computational time.

Keywords— Soft Computing, Fuzzy Inductive Reasoning, Entropy-based Feature Selection, Nearest Neighbours Selection, Energy Modelling

I. INTRODUCTION

Digital transformation, disruptive technologies, flexibility markets and new business models are bringing a new era into the energy sector. The digitalization of the medium and low voltage [1] have unlock the Smart Grid's potential with demand response mechanisms, online customer interactions, dynamic electricity tariff and dynamic load balancing.

In addition to that, the inclusion of renewable energy production in the medium and low voltage, the increase of power quality assurance and the good omen for home batteries requires new solutions and technologies to cope with them. In fact, most of the features and proposed scenarios are based on reliable, accurate and fast energy predictions.

In the recent years, the importance of load forecasting extends also to buildings and homes; detection of potential demand response programs [2], peaks that may increase the energy bill in a dynamic tariff framework [3] or the use of predictions to create messages to change the energy behaviour of the tenants.

Although huge improvements have been achieved in forecasting accuracy with Soft Computing and Machine Learning techniques, there is still a big challenge in the model parametrization, e.g. in soft computing techniques, where force tasks are not as efficient as in deep learning. In fact, to know how to tune parameters can lead to a good or bad performance in forecasting applications, smart grid simulations and control scenarios.

For instance, determination of the most suitable number (k) of Nearest Neighbours (kNN) for a better prediction is a major difficulty in models that use this algorithm. Smaller k brings higher noise sensitivity, whilst larger k causes smoother decision boundaries and lower noise sensitivity [4]. To overcome these issues, in some methods, k is tuned in order to find the optimal mapping function, but it is a time-consuming process [5].

In this paper, we are addressing the importance of the parameter selection of k in kNN, in a Soft Computing technique called Fuzzy Inductive Reasoning (FIR) [6][7]. Although its popularity is not comparable to other Soft Computing techniques such as Neural Networks (NN), this methodology has been proved to model real complex systems with high accuracy compared to other Artificial Intelligence (AI) and statistical techniques[8][9][10][11].

Model parametrization in FIR has led to an important number of publications [12][13]. Most of the parametrization approaches in the literature are performed from an offline perspective and take considerable computational time and resources. In addition to that, none of them is related to the optimal kNN parameter selection, which is key during the forecasting process and accuracy of predictions.

In this article, we perform an analysis of the impact that kNN has in FIR prediction and we present an improved version that uses a kNN optimal selection algorithm during the FIR prediction phase. To this end, a real smart grid forecasting application, i.e. electricity load forecasting, has been chosen in this study. However, notice that the enhanced FIR methodology presented in this paper can be applied to any other application area.

The paper is structure as follows: in section II, the Standard and Flexible FIR methodologies are summarized and the challenge of kNN selection is identified. Then, section III presents a new approach to select the most optimal number of nearest neighbours during FIR prediction phase, called K nearest neighbour Optimal Selection (KOS). Next, section IV presents the datasets used, which come from three different buildings belonging to the UPC (Universitat Politècnica de Catalunya), the experiments performed and the discussion of the results encountered when KOS is used to the datasets. Finally, section V points out the conclusions of our research and the near future work.

II. FUZZY INDUCTIVE REASONING

A. Standard and Flexible FIR

The conceptualization of the FIR methodology arises out of the General System Problem Solving (GSPS) approach proposed by Klir [14]. This methodology of modelling and simulation has the ability to describe systems that cannot be easily described by classical mathematics or statistics, i.e. systems for which the underlying physical laws are not well-understood [6]. A FIR model is a qualitative non-parametric

model based on fuzzy logic. The FIR model consists of its structure (relevant variables or selected features) and a pattern rule base (a set of input/output relations or history behaviour) that are defined as if-then rules.
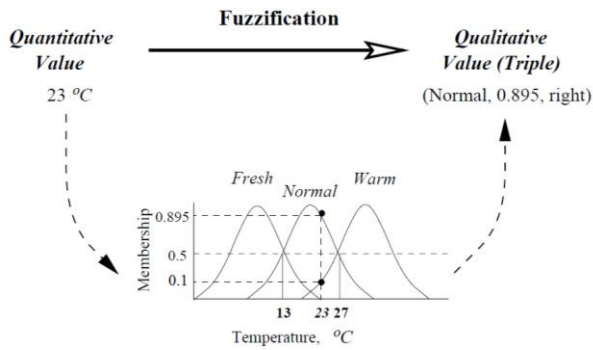

Fig. 1. Fuzzy inductive reasoning (FIR) fuzzification of a temperature value of 23ºC

Before starting the process of finding a FIR model in order to make predictions from the data, it is necessary to fuzzify[1] the data so that the search space is reduced and the optimization process is speeded up. FIR embraces a slightly different approach than other fuzzy techniques to solve the uniqueness problem. Rather than mapping into multiple fuzzy rules, FIR only maps into a single rule with the largest likelihood. To this end, FIR converts each quantitative value into a qualitative fuzzy triple, i.e. the *class*, the *membership* and the *side* values. The class value represents a discretization of the original real-valued variable. The fuzzy membership value denotes the level of confidence, expressed in the class value chosen to represent a particular quantitative value. The side value indicates whether the data point is to the left or the right of the peak of the corresponding fuzzy membership function. The side value, which is a specialty of the FIR technique since it is not commonly used in fuzzy logic, is responsible for preserving, in the qualitative triple, the complete knowledge that had been included in the original quantitative value. And it is thus possible to regenerate the quantitative value precisely. Fig. 1 illustrates the process of fuzzification by means of an example. Hence, a temperature of 23∘C would be fuzzified into the class normal with a side value of right and a fuzzy membership value of 0.895.

Notice that FIR is performing a mapping to a single rule, the one with the largest likelihood: (Normal, 0.895), while usually fuzzy logic approaches perform a mapping into multiple fuzzy rules, in this example: (Normal, 0.895) and (Warm, 0.1). For a deeper insight into the fuzzification process refer to [6].

The process of obtaining a FIR model structure corresponds to a Feature Selection Process (FSP), as it was explained in [8]. Table 1 presents an example of mask, i.e. structure that contains the selected more relevant features. Each negative element in the mask is called mask input (*m*-input). The single positive value denotes the mask output (*m*-output). It exhibits a causal relation with the output, i.e. it influences the output up to a certain degree. In the example of Table 1, the prediction of the output at the current time, $O(t)$, is directly related to the variables I_1, I_2, I_3 and O in different

times t, i.e. $I_3(t-2\delta t)$, $I_2(t-\delta t)$, $O(t-3\delta t)$ and $I_1(t-\delta t)$. δt represents a time span into the past.

Table 1. Example of mask for a system with three inputs (I_1, I_2, I_3), one output (O) variables and complexity equal to 5

t \ x	I_1	I_2	I_3	O
t - 2δt	0	0	-1	0
t - δt	0	-2	0	-3
t	-4	0	0	+1

The optimal mask function of FIR is used to obtain the best mask, i.e. the best FIR structure, for the system under study. The procedure consists in finding the mask that best represents the system by computing a quality measure for all possible masks, and selecting the one with the highest quality. An exhaustive explanation, analysis and examples of this procedure can be found in [6][1][7].
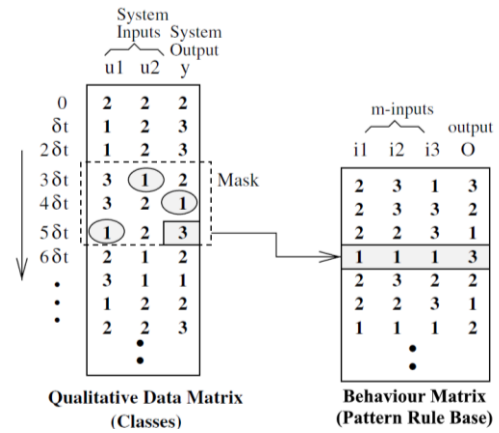

Fig. 2. FIR pattern rule base construction process diagram (with an example containing two inputs and one output)

Once the most relevant features are identified (feature selection), they can be used in any modelling methodology. As it is shown in Fig. 2, the optimal mask is used to obtain the set of pattern if-then rules (called behaviour matrix) from the fuzzified training data set [8].

The left side of Fig. 2 shows an excerpt of the training data already fuzzified (only the class values are shown). The mask is shifted downwards along the class training data. The round shaded ''holes'' in the mask denote the positions of the *m*-inputs, whereas the square shaded ''hole'' indicates the position of the *m*-output. The class values are read out and placed next to each other in the behaviour matrix that is shown on the right side of the Fig. 2. The shaded rule of this figure can be read as follows: ''If all the *m*-input (i1, i2 and i3) have values of '1' (corresponding to 'low') then the output, o, assumes a value of '3' (corresponding to 'high'). FIR is able to infer the model of the system under study very quickly; it is a good option for real time forecasting and is able to deal with missing data as it has been already proved in a large number of applications [10]. However, its capacity to deal with missing data decreases significantly when the complexity of the mask is big, because it implies the generation of a big number of pattern rules in the behaviour matrix containing Missing Values (MVs).

---

[1] Process of converting the quantitative value into a qualitative fuzzy triple, i.e. the class, the membership and the side values.

Recently, an improved version of Standard FIR called Flexible FIR Prediction has been demonstrated [15], which can cope with missing information in the input pattern as well as learn from instances with missing values in the behaviour matrix. An extended study of this improved version of FIR can be found in [16]. Moreover, comparisons of Standard, Flexible FIR and other statistical and AI techniques have been performed in [8][9][15].

B. Fuzzy Forecasting Process and Number of Nearest Neighbours

Once the behaviour matrix and the mask are available, a prediction of future output states of the system can take place using the FIR inference engine, as described in Fig. 3. This process is called qualitative simulation. The FIR inference engine is based on the kNN approach, commonly used in the pattern recognition field. The forecast of the output variable is obtained by means of composition of the potential conclusion, which results from firing the kNN rules whose antecedents have best matching with the actual state.
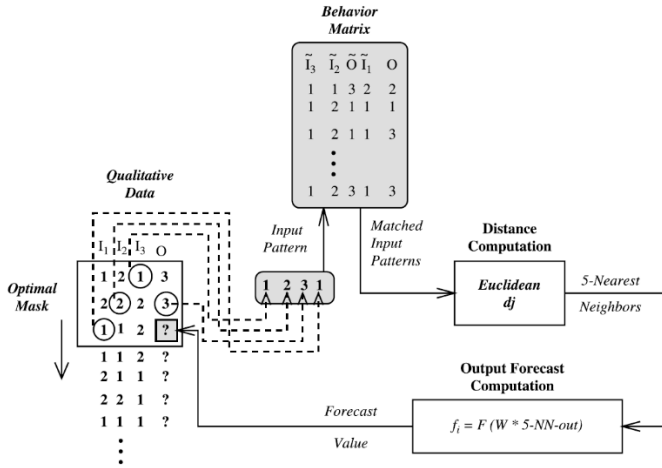

Fig. 3. Qualitative simulation process diagram (with an example containing three inputs and one output)

As can be seen in the left hand side of Fig. 3, the mask is placed on top of the qualitative data matrix (fuzzified test set), in such a way that the output matches with the first element to be predicted. The values of the inputs are read out from the mask and the behaviour matrix (pattern rule base) is used, as it is explained latter, to determine the future value of the output, which can then be copied back into the qualitative data matrix. The mask is then shifted down one position to predict the next output value. This process is repeated until all the desired values have been forecast.

The fuzzy forecasting process works as follows: the input pattern of the new input state is compared with those of all previous recordings of the same input state contained in the behaviour matrix. For this purpose, a normalization function is computed for every element of the new input state and an Euclidean distance formula is used to select the kNN, the ones with smallest distance, which are used to forecast the new output state [6]. The contribution of each neighbour to the estimation of the prediction of the new output state is a function of its proximity. This is expressed by giving a distance weight to each neighbour, as shown in Fig. 3. The new output state values can be computed as a weighted sum of the output states of the previously observed kNN [5].

In studies and applications about FIR, the new output state is, typically computed using five nearest neighbour [10]. It has been always considered, based on the results, that five nearest neighbours were a good compromise between dispersion and accuracy. However, this assumption is far from an optimal solution. As pointed out in section IV, including one more/less neighbour can affect significantly in the performance of the FIR prediction. In the next section, we deal with this problem.

III. KNN OPTIMAL SELECTION

The enhancement proposed in this paper is to develop an algorithm that selects an optimal parameter for the k nearest neighbours to be used with Flexible FIR Prediction, which is called Knn Optimal Selection (KOS). As it was previously explained, the new output state is determined directly from the k nearest neighbours selected. The KOS algorithm is depicted in Fig. 4.
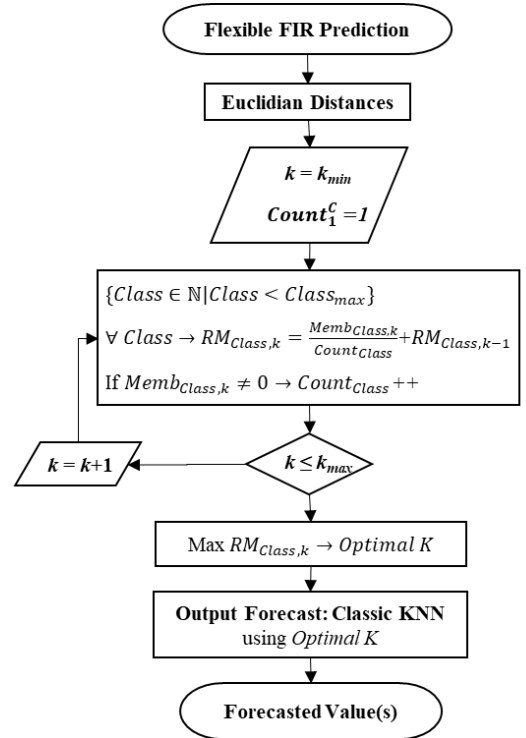

Fig. 4. Flow diagram of the KOS algorithm. RM stands for Relative Membership matrix

The Euclidean distance between every pattern found in the behaviour matrix with the same input pattern (neighbours) and the new input pattern to which we want to predict its output, is computed. Afterwards, the neighbours are sorted with respect their distance, from the smallest to the highest distance.

The parameter k is initialized with a k_{min} and the counter parameters ($count_i$) are initialized with a value of 1. There are as many $count$ parameters as possible output classes c.

Then, the Relative Membership matrix (RM) is computed. The matrix has as many rows as output classes and k_{max}-k_{min} columns, and it works as follows: we take the output class ($Class_k$) and membership ($Memb_{Class,k}$) from the first neighbour, i.e. the one with the minimum distance ($k=k_{min}$) and we placed it in the position $RM_{Class,k}$. For each neighbour, only one class contribution to the RM matrix can be added, therefore, the rest RMs with this k will be 0. If the next

neighbour (k+1) has as output a class already observed, the membership value for this class is divided by the number of neighbours already observed in the class and added up to the previous values of the RM in that class ($RM_{Class,k-1}$).

Once all the neighbour have been treated, i.e. have been incorporated into the RM matrix, the optimal number of k nearest neighbours is selected as the maximum membership function. If there are several k that has the maximum value, we pick the lower k out of it. The selected optimal k is used by the FIR methodology to compute the output forecast, using the qualitative simulation process described in section II.*B*.

The idea behind the KOS approach is to perform a kind of membership aggregation function of all the neighbours with respect their belonging to the output classes. The output class that has the higher aggregated value for that specific input pattern is the one that represents better the behaviour of the system. Then, once the output class is selected, the optimal k is the minimum k that has the highest aggregated membership value.

Table 2 presents an example of a RM that helps to clarify the KOS process. The matrix contains 3 different classes (1, 2, 3) and 6 neighbours, from k_{min} (1st neighbour) to k_{max} (6th neighbour). The first neighbour output class and membership are 1 and 0.7, respectively. Thus, $RM_{1,1}$=0.7, $RM_{2,1}$=0 and $RM_{3,1}$=0. The counter for class 1 increases ($Count_1$=1+1), while the rest counters remain the same ($Count_2$=1 and $Count_3$=1). The next closest neighbour output class is 1 as well. With a membership of 0.9, the value of $RM_{1,2}$ will be $RM_{1,1}$+(0.9/$Count_1$). When the RM matrix is completed, the *Optimal K*, is k=5. Notice that we have chosen 5 instead of 6 because we select the first maximum RM value, i.e. the lowest k with the maximum RM value.

Table 2. Example of a RM matrix with 3 different classes (1, 2, 3) and 6 neighbours; from k_{min} (1st neighbour), to k_{max} (6th neighbour)

C_N \ k	1	2	3	4	**5**	6
1	0+0.7	$0.7+\frac{0.9}{2}$ = 1.15	1.15	1.15	$\mathbf{1.15+\frac{0.62}{3}}$ = **1.36**	1.36
2	0	0	0.8	$0.8+\frac{0.75}{2}$ =1.17	1.17	1.17
3	0	0	0	0	0	0.95

IV. EXPERIMENTS AND RESULTS

A. Dataset

Similarly to [16], data of 3 buildings of the Universitat Politècnica de Catalunya (UPC) was obtained for this study, in order to have a training/test sample with high diversity of consumptions. They have different profiles of usage (teaching, library and administration building), belong to three different campuses and are located in different cities. Thus, affecting different climatology (temperature, humidity, solar radiation, etc.), consumption patterns, schedules and working days. The buildings included are: 1) one administrative building (Edifici Campus Terrassa) in ETSEIAT faculty in Terrassa; 2) the Library of EPSEVG faculty in Vilanova; 3) Building C6 with different classrooms at FIB faculty in Barcelona. The energy consumptions of these 3 buildings have been collected through a remote metering system every hour. Therefore, there are 24 recordings per day and per location over one year, which means 70080 hourly consumptions used in the experiments.

For all the three buildings, the dataset comprises a whole year of electricity consumption, from 13/11/2013 to 12/11/2014, from which 91% was separated for training and the remaining 9% for testing. The testing data comprises 35 different days (i.e. 35 test sets) distributed equally through the whole year; meaning around 9 days per season and taking into account the seven days of the week (from Monday to Sunday). The 9% of data used for testing (only in historical consumptions output variable) is substituted by MVs. By choosing these days, we aim to evaluate the models against the changes caused by seasonal period(s) and day of the week.

B. Experiments

The aim is to evaluate if the KOS is able to perform, on average, as good as the number of nearest neighbours that performs better, and to understand the implications of the parameter k in FIR.

To do so, as it is shown in Fig. 4, we perform three different experiments: i) with k from 1 to 15 neighbours, ii) with the new KOS and iii) with a random value of k. Thus, we perform 24h Flexible FIR predictions with a selection of 17 (15(i) + 1(ii)+ 1(iii)) different nearest neighbours, in the Flexible FIR forecasting process.

In all the experiments, first, the training test is used to compute the 4 more relevant past consumptions based on FIR Feature Selection Process (FSP). Right after, the process of FIR model creation uses the 4 most important past consumptions, plus two more input variables: if it is a working day and the hour of the day. Notice that FIR FSP is computed only once.

In i), when Flexible FIR forecasting is performed, we select the parameter k equal to 1. After that, the prediction error is computed based on the 35 test days. Then, k is increased 1 unit, and again we perform the prediction error for the 35 test days. The experiment is repeated until k is equal to 15. This is done for each single prediction, i.e. for each hour. Notice that in this case, k is the same for every hourly prediction, in the 35 test days.
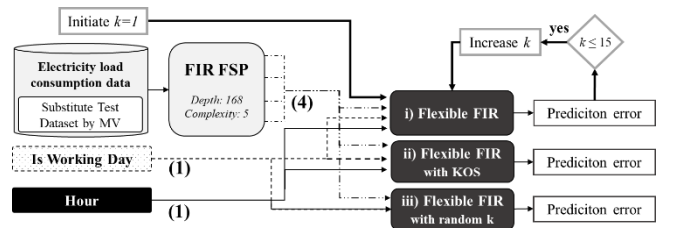

Fig. 5. Scheme of the whole experiment. The numbers in parentheses stand for the number of variables added in the model. In Flexible FIR with KOS, k can adopt a value in the range [1, 15]

In parallel, it is performed the same experiment but with the KOS algorithm (ii). The KOS selects, in each hourly prediction, the most appropriate k among the first 15 nearest neighbours, computing the formula explained in section III. Therefore, the computational cost is much lower than in (i).

Finally, a last experiment (iii) is performed selecting a random k from 1 to 15 in each hourly prediction.

C. Model Parameters

In this study, the modelling process consists of: 1) FSP 2) use of the relevant features to derive a FIR model, 3) selection of the flexible FIR strategy in case of missing data and 4) the minimum and maximum k nearest neighbours to be used in the three experiments.

The FSP is applied only to the historical consumption data and not to the hourly and daily information, as shown in Fig. 5. It is decided to follow this strategy because the hourly and daily information contains only the hour of the day and if it is or not a working day, respectively. Therefore, the valuable information is gained with the actual value not with the previous ones. However, previous consumptions contain information patterns from where important knowledge could be extracted.

As it is explained in [16], the selection of the depth and number of variables is a crucial issue that can affect those methods that are more sensitive to the curse of dimensionality. It has been empirically determined, for electrical load consumption applications, that more than four variables and depths higher than 72 hours, do not increase significantly the quality of the FSP of FIR, whereas computational cost (in terms of time) does exponentially [15].

In [8], we have demonstrated that for these data the optimal configurations are when mask depth is 24 + 24: previous 24 hours and the past 24 hours of the previous week (48 past values in total that corresponds to a depth of the mask of 168). With regard to the complexity of the mask, the four past consumptions selected by FIR are taken into account in all the models and the mask increases to five and six when the variables working day and hour are added. Regarding the fuzzification parameters, three classes and the equal frequency partition algorithm have been used to discretize the electrical load consumption and hour of the day variable. Working day variable is binary and, therefore, it has been discretized into two classes. Concerning the flexible FIR strategy, we have opted for the Classic FIR KNN (aKnn) [16], to simplify computations and because our aim is not to evaluate Flexible FIR strategies but the KOS algorithm. Finally, for the first branch of experiments (i), k goes from 1 to 15. For the KOS (ii) and random (iii) approaches, k can adopt any value from k_{min} equal to 1, to k_{max} equal to 15.

D. Evaluation Criteria

There are many measures of forecast's accuracy in the literature [17]. We require a statistical quality measure, which is able to compare the different forecasting methods in buildings with different average loads.

The Mean Squared Error is not suitable to evaluate the performance of the model when values in the datasets differ in magnitude. For example, values predicted in some buildings are in the order of 132 kWh, whereas, in others is in the order of few kWh. The current research also considered to use the Mean Absolute Percentage Error (MAPE) to offer a forecasting performance from a multi-dimensional perspective. However, MAPE puts a heavier penalty in negative errors than in positive errors [16][17].

This observation led to the use of the so-called "symmetric" Mean Absolute Percentage Error (sMAPE) [17] defined by equation 1.

$$sMAPE = 200 * \frac{1}{N}\sum_{t=1}^{N} |(y_r(t) - y_f(t))|/(y_r(t) + y_f(t)) \quad (1)$$

where $y_r(t)$ is the real output value, $y_f(t)$ is the forecasted output value and N is the size of the test data set.

In addition, it has to be highlighted that measures based on percentage errors have the disadvantage of being undefined when $y(t) = 0$ for any t in the period of interest, and having an extremely skewed distribution for values of $y(t)$ very close to zero. In our experiments, it has been verified that none of the consumption data points $y(t)$ are equal or very close to zero.

E. Results

Fig. 6, Fig. 7, and Fig. 8 show the error evolution computed with sMAPE in different Flexible FIR configurations. The percentage for each k is an average of the error prediction in the 35 test days considering the three buildings of study. The values of k, from 1 to 15 are the results of Flexible FIR when the output forecast computation uses k nearest neighbours (i). k equal to 16 represents the experiment with the new KOS in each prediction (ii), while k equal to 17 represents the result with a random k in each prediction (iii).

On the one hand, the results of Building C6 shows that the performance of KOS is slightly worse than the three best number (k) of neighbours, i.e. 6, 7 and 8 but very close, with less than 0.4% of difference.
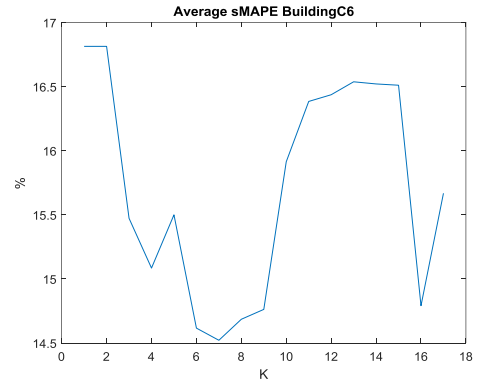

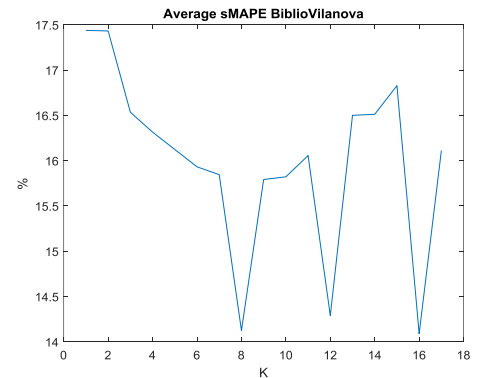Fig. 6. sMAPE error evolution against different number of neigbhours for Building C6

Fig. 7. sMAPE error evolution against different number of neigbhours for Vilanova Library

On the other hand, the results in Vilanova Library and the Administrative Building in Terrassa show that the performance with KOS is slightly better or almost equal to the optimal number of neighbours. Fig. 9 shows the error evolution of the three buildings aggregated.
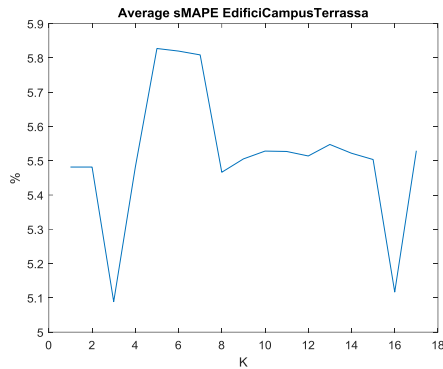
Fig. 8. sMAPE error evolution against different number of neigbhours for an Administrative Building in Terrassa

As it can be observed, on average, the configuration with the best results is when k is 16, i.e. when the KOS is used.
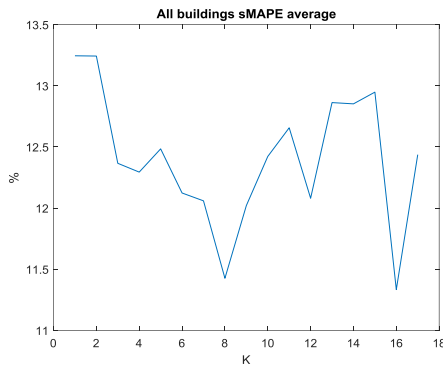

Fig. 9. sMAPE average of the three buildings used in our experiments

V. CONCLUSIONS AND FUTURE WORK

How would you choose the number of k nearest neighbours in your Standard or Flexible FIR model, if you were aware that this number affects to the forecasting accuracy? Would you choose a random k? Would you run n different models and choose k based on the best result? In this article, we introduce the KOS that works with Standard and Flexible FIR, which is able to decide in each new prediction, the optimal parameter of k. The results obtained point out that the number of neighbours selected in Standard and Flexible FIR affects significantly to the performance of the prediction. On average, there is more than two percentage points of difference between k_{16} and k_1. This difference is higher in buildings such as *BiblioVilanova,* with more than three percentage points between the worst and best prediction. Additionally, there is not a clear pattern in the error evolution, as it can be seen in Fig. 6, Fig. 7, Fig. 8 and Fig. 9, thus, the use of KOS is a clear advantage in these scenarios. The outcome of our study sheds light on robust Soft Computing methodologies for smart home, smart buildings and smart grid applications. It has been proved that, on average, the forecasting accuracy of Flexible FIR combined with KOS improves. Our algorithm helps to decide in each hourly prediction, which is the optimal number of neighbours to compute the next output state in Standard and Flexible FIR. As future work, we are going to extend the experiments to more UPC's buildings, we will investigate the impact of higher k_{max} and its integration in a FIR Type-2. We also plan to integrate Flexible FIR and KOS in a smart home Internet of Things (IoT) platform that will provide home context awareness such as prediction of events or patterns detection.

VI. REFERENCES

[1] European Comission. Benchmarking smart metering deployment in the EU-27 with a focus on electricity [com/2014/0356 final]. Technical report.

[2] J. Justin, M. Vijaya, R. K Uma, "Impact of demand response contracts on short-term load forecasting in smart grid using SVR optimized by GA", *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Vellore, 2017, pp. 1-9.

[3] E. Onur, T. Akın, I. A. Tahir, S. Uğur S. 2017. "Implementation of a dynamic energy management system using real time pricing and local renewable energy generation forecasts", *Energy*, 2017, vol. 134(C), pp. 206-220.

[4] K. Beyer, J. Goldstein, R. Ramakrishnan and U. Shaft, "When is nearest neighbormeaningful?" *International Conference on Database Theory*, 1999, pp. 217–235.

[5] D.J. Hand and V. Vinciotti, "Choosing k for two-class nearest neighbour classifiers with unbalanced classes", *Pattern Recognition. Letters,* 2003, vol. 24(9), pp. 1555–1562.

[6] A. Nebot, and F. Mugica, "Fuzzy Inductive Reasoning: a consolidated approach to data-driven construction of complex dynamical systems", *International Journal of General Systems*, 2012, vol. 41(7), pp. 645-665.

[7] A, Escobet, A. Nebot, F.E. Cellier, "Visual-FIR: A tool for model identification and prediction of dynamical complex systems", *Simulation Modelling Practice and Theory*, 2008, vol. 16, pp. 76-92.

[8] S. Jurado, A. Nebot, F. Mugica, "Hybrid methodologies for electricity load forecasting: Entropy-Based Feature Selection with Machine Learning and Soft Computing Techniques", *Energy*, 2015, vol. 86, pp 276-291.

[9] S. Jurado, J. Peralta, A. Nebot, F. Mugica, and P. Cortez, "Short-term electric load forecasting using computational intelligence methods", *FUZZ-IEEE 2013: 2013 IEEE International Conference on Fuzzy Systems*, 2013, Hyderabad, 2013, pp. 1-8.

[10] A. Nebot, F. Mugica, F. Cellier, M. Vallverdú, "Modeling and Simulation of the Central Nervous System Control with Generic Fuzzy Models", *Transactions of The Society for Modeling and Simulation*, 2003, vol. 79(11), pp. 648-669.

[11] P. Gómez, A. Nebot, S. Ribeiro, R. Alquézar, F. Mugica and F. Wotawa, "Local maximum ozone concentration prediction using soft computing methodologies", *Systems Analysis Modelling Simulation*, 2003, vol. 43(8), pp. 1011–1031.

[12] J. Acosta, A. Nebot, P. Villar, J.M. Fuertes, "Optimization of fuzzy partitions for inductive reasoning using genetic algorithms", 2007 *International Journal of Systems Science*. vol. 38. pp. 991-1011

[13] I. Paz, A. Nebot, E. Romero, F. Mugica, A. Vellido, "A methodological approach for algorithmic composition systems' parameter spaces aesthetic exploration" *IEEE Congress on Evolutionary Computation (CEC)*, 2016, pp. 1317-1323.

[14] J. Klir, D. Elias, *Architecture of Systems Problem Solving*, 2nd. Ed., Plenum Press, New York, 2002.

[15] S. Jurado, A. Nebot, and F. Mugica, "A flexible fuzzy inductive reasoning approach for load modelling able to cope with missing data", *8th International Conference on Simulation Tools and Techniques (SIMUTools '15)*, 2015, pp. 349-356.

[16] S. Jurado, A. Nebot, F. Mugica and M. Mihaylov, "Fuzzy inductive reasoning forecasting strategies able to cope with missing data: A smart grid application", *Applied Soft Computing,* 2017, vol. 51, pp. 225-238.

[17] R.J. Hyndman, A.B. Koehler, "Another look at measures of forecast accuracy", *International Journal of Forecasting*, 2006, vol. 22, pp. 679-688.

[18] S. Makridakis, "Accuracy measures: Theoretical and practical concerns", *International Journal of Forecasting*, 1993, vol. 9, pp. 527 – 529