Computational Intelligence in Forecasting – The Results of the Time Series Forecasting Competition

Martin Štěpnička, Michal Burda
Institute for Research and Applications of Fuzzy Modeling
Centre of Excellence IT4Innovations, Division University of Ostrava
30. dubna 22, 701 03 Ostrava, Czech Republic
e-mail: martin.stepnicka@osu.cz, michal.burda@osu.cz

***Abstract*—The aim of this paper is to present the results of the time series forecasting competition that was organized within the IFSA-EUSFLAT 2015 conference.**

I. Introduction

This paper presents the results and some observations from the international time series forecasting competition entitled the *Computational Intelligence in Forecasting* (abbr. CIF) which was organized within the IFSA-EUSFLAT 2015 conference. Therefore, this article does not bring a detailed description of a newly proposed single method and its experimental evaluation, as it is usually provided in, say, standard articles on computational intelligence (abbr. CI) methods for time series forecasting.

On the other hand, this article brings an analysis of the forecasting performance of several methods, including descriptions, comments, observations, and other findings that can be obtained only from an independent evaluation. This added value, which may never be brought by authors itself in their own articles, is something that we find very valuable and desirable.

No author demonstrates weaknesses of his/her forecasting methods. This drawback has an obvious reason and it is not surprising, as negative results are often viewed as no-results. From such articles, it is very difficult or even impossible to learn, which method should be used in which situations, what are the benefits of using one or any other, and what are the methods (or steps in tuning the methods) one should rather avoid. This article is an attempt to address such natural questions and although we do not dare to say that we answer them, we are convinced that the content of this article is of an unquestionable merit.

A. Time Series Forecasting Competitions, Fuzzy Methods, "The Others"

The meaning of the quotes in the title of this section is twofold. Firstly, by "the others" we mean the other "competitions" apart from the time series forecasting ones, and at the same time, we mean the other "methods" apart from the fuzzy ones. The explanation is as follows.

Organizing competitions in many areas of application of applied mathematics and computer science became a standard way of getting an independent evaluation and independent information on a huge number of methods that are being proposed. Competitions show their potential in pushing research forward due to the creative approach of organizers who define new and up-to-date tasks for participants. Moreover, the results obtained by the independent evaluators, including the ones with poor results, are viewed as the best source of information for further improvements for the contestants themselves and thus even non-successes have to be positively accepted as valuable information.

Within the CI community, the competitions became very popular in some areas. Note that the data-mining competitions focused on classification tasks attract enormous numbers of contestants, for example the annual competition at the KDD conference attracted over 1000 of them.

Coming closer to the main areas of the IEEE CIS, we may observe that competitions also take an important part of the key conference. However, competitions are organized rather in the neural network and mainly in the evolutionary computation area. Recall that e.g. 4 competitions were organized within the IEEE WCCI 2014, but 1 of them was located on the IEEE IJCNN part of the congress and the other 3 competitions on IEEE CEC, none on FUZZ-IEEE. Note also that the upcoming IEEE CEC web page announced already 7 accepted competitions in the optimization area. Of course, although rarely, there are also competitions being organized in the area of fuzzy modeling but mainly focusing on fuzzy control, for example on FUZZ-IEEE 2009 or FUZZ-IEEE 2013. The time series forecasting competition attracting "fuzzy community" has not been opened yet, although, as we show below, this area is more and more attractive for people dealing with fuzzy methods.

The history of rather big time series forecasting competitions dates back to 1980's. In particular, the first very really broad competition was organized by S. Makridakis et al. [1] already in 1982 and was called the **M-Competition**. One of the interesting outputs of that competition was that statistically sophisticated or complex methods do not necessarily provide more accurate forecasts than simpler ones. This output was not very positively accepted by the statisticians in the forecasting community, which is not surprising, as the community developing more and more complex methods is on the other side of the river. One of the articles aiming at weaknesses of such a competition was even entitled "The competition to end all competitions", see [2].

Therefore, the organizers prepared another competition entitled the **M2-Competition**. The results of the competition confirmed the outputs of the previous competition, see [3].

Finally, on the edge of the millennium, the organizers prepared the last competition entitled the **M3-Competition**, which consisted in forecasting of 3003 time series from distinct fields and of distinct frequencies (yearly, quarterly, monthly etc.). The conclusions of this competition were again similar to those of the previous ones [4], [5]. Nowadays, the data-sets of time series used in all the three **M**-competitions is a generally accepted benchmark provided by the *International Institute of Forecasters* as the professional authority in the field of forecasting.

Focusing on CI methods, there was only a single participant (using a forecasting method based on neural networks) from this area that took part in **M3-Competition** and provided rather weak performance. Of course, such a result generally does not mean anything bad about neural networks and their potential, as much more it gives a zero information due to the lack of methods and models (even a powerful method may lead to a weak performance when the particular models are not well-tuned).

In order to extend the limited attention of the neural network (abbr. NN) and generally the whole CI society, Sven Crone organized a competition entitled **NN3**, which was open also to statistical methods, but they were evaluated only as a benchmark, while only CI methods were competing for victory. The positive impact on closing the competition to CI methods led to a very high number of submissions, in particular to 59 submissions from the CI area as well as statistical benchmarks. The fact that CI methods may compete with classical statistical methods is among the main findings mentioned in [6], especially ensembles of CI methods performed better than combinations of statistical methods. On the other hand, it should be mentioned that the vast majority of the participants (methods) were from the NN area and then from other CI areas such as regression tress, evolutionary algorithms, or hybrid systems. Up to our best knowledge, there were only two participants using some fuzzy approaches to the forecasting task. Out of these two, one of them used specific fuzzy neural network [7] and thus, the only fuzzy approach not employing any fragments of NN or other areas of CI was the one employing Fuzzy transform [8] and fuzzy rules with a specific inference mechanism called the Perception-based Logical deduction [9], which later on led to the linguistic approach described e.g. in [10], [11].

B. Fuzzy Methods for Time Series Forecasting

The interest of the "fuzzy" community into the time series forecasting is unquestionably increasing and it is by far not negligible. Note that the number of articles in the Web of Science that contain the conjunction of the words "fuzzy" and "time series" directly in the title increases rapidly, it was 50 articles in 1990's, then 319 in 2000-2009 and finally, it is 244 since 2010 which is again a significant increase having in mind, that still many articles published in 2014 are not yet included in the database.

This huge interest includes many articles not only aiming at forecasting of time series, but also at their linguistic description [12], [13] or their comparison and clustering [14], [15]. However, unquestionably most important is still their forecasting. Among the distinct fuzzy methods aiming at this task, we may identify several main directions of the research. For example, a study presenting Takagi-Sugeno rules in the view of the Box-Jenkins methodology [16] or the works dealing with the linguistic approach [11], [17] were published. In most cases, various neuro-fuzzy approaches, which lie on the border between neural networks, Takagi-Sugeno models, and evolving fuzzy systems, are very often successfully used [18]–[20].

A notable number of articles is also often published on the so called "fuzzy time series". This term has been introduced by Song and Chissom [21] and attracted many followers. There are two issues related to the vast majority of the articles dealing with this notion. The first one consists in mathematically incorrect definitions of the fundamental notions including the definition of the fuzzy time series itself. The second relates to incorrect or inappropriate experimental evaluation of such results. Mainly the results are experimentally evaluated on say one or two time series, which does not bring any information on the forecasting potential, but there are more serious flaws. For example, most of the articles build a time series forecasting model on some data and then the model is used to predict the same data, which makes the experimental results and comparisons to other methods useless and brings zero information, see e.g. [22], [23]. There are papers that correctly make a distinction between the training set and the testing set, for example [24], but often the prediction looks like time-delayed testing data which gives rise to a suspicion that the testing data are not used to determine the model, however, they are "feeded" to a one-step-ahead forecasting model instead of forecasting from the forecasted values, which again would mean that the experimental evaluation would not bring any valuable information. This generally does not mean that the models based on the fuzzy time series concept may not poses any forecasting power. The main idea of describing the time series values by fuzzy rules is indeed meaningful. However, an independent and correct experimental evaluation is highly desirable.

All in all, the competition to distinct data processing tasks makes sense because of many reasons, and although they are usually negatively accepted at the beginning of their existence, finally they become a standard tool for getting valuable independent information. The fact that this "competition culture" is not yet that common in the fuzzy community is only a reason to introduce it, not to neglect it. Especially the time series competitions are so far neglected in the fuzzy community. On one hand, it is somehow natural as even in the statistical time series forecasting community were these competitions accepted only with some difficulties at the edge of the millennium and the CI community (mainly NN community) overtook this valuable tool, say 10 years ago. On the other hand, in data mining, stochastic optimization, and many other areas, this is a common part of activity bringing another insight into the state-of-art and having in mind a huge explosion of articles (and consequently methods and models) on fuzzy approaches to time series forecasting, the lack of independent evaluation cannot be ignored anymore.

II. The Setting of the Competition

A. Rules of the Competition

The objective of the CIF competition was to forecast a data-set of given time series from distinct domains as

accurately as possible using methods from the computational intelligence area while applying a consistent methodology. The data consisted of time series with different time frequencies, that include yearly, quarterly, monthly, and daily data.

The competition was open to all methods of computational intelligence, including fuzzy method, artificial neural networks, evolutionary algorithms, decision and regression tress, support vector machines, hybrid approaches etc. used in all areas of forecasting, prediction and time series analysis, etc. Ensemble techniques were also allowed, if they employ any computational intelligence method.

The only evaluation criterion was the Symmetric Mean Absolute Percentage Error (SMAPE):

$$\mathrm{SMAPE} = \frac{1}{h}\sum_{t=1}^{h}\frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}, \tag{1}$$

where F_t is t-th forecasted value, A_t is t-th actual (real) value, and h is the forecast horizon. A method with the lowest average SMAPE over all time series is considered as a winner.

B. The Competition Data-set

The competition data-set consisted of 91 time series from banking, web hosting, and health domains. The frequencies of the time series (i.e. the frequencies at which the measures were reported) were daily, monthly, quarterly, and yearly. The lengths of the time series differed significantly, with shortest time series being that of yearly frequency. The forecasting horizons (i.e. the number of future values to be forecasted) was set by taking into account the length of the time series and its frequency: horizon differs from 4 for short yearly time series to 180 for quite long daily time series. See Table I for more details.

C. Benchmarks

As a benchmark, several well-known (non-fuzzy) forecasting methods were evaluated on the competition time series. Namely, we have chosen ARIMA, Exponential Smoothing, the Theta method, and Random Walk; for details on these methods see e.g. [25]–[27]. All these methods were executed with default settings as provided by the *forecast* package [28] of the R statistical environment [29]. The reason was to use a standard tool that provides us with an automatic model selection and parameter tuning, and thus, to avoid any bias from a naive implementation of the above listed methods. Due to the use of the R environment, we denote the methods as R-ARIMA, R-ETS, R-Theta, and R-RW, respectively.

Also two ensembling techniques were added as a benchmark. Both of them aggregate the forecasts of the four methods listed above. First ensemble is a simple average of the forecasts (R-AVG), second is the Fuzzy Rule Based Ensemble (R-FRBE) [30], [31] that is implemented in the *lfl* package [32], [33] of the R statistical environment [29].

As of version 1.0 of the *lfl* package, FRBE is a technique that produces the forecast based on an ensemble of R-ARIMA, R-ETS, R-Theta, and R-RW. The ensemble is computed as a weighted average of the individual forecasts. The weights are inferred from some characteristics computed from the time series being forecasted. The following time series characteristics are involved: frequency, length, skewness, kurtosis, trend strength, season strength, variation coefficient, and stationarity.

To estimate weights of the individual methods, a pre-defined rule base of fuzzy association rules is used together with a so-called *Perception-based Logical Deduction* (abbr. PbLD). PbLD is a specific inference method that is suitable to be used with rule bases constructed from linguistic expressions that are well modelled with fuzzy sets. PbLD assumes that the rules can be partially ordered by their *specificity* and the inference is influenced by that order: only the most specific rules are selected for the inference [9], [34], [35]. Note that this method is closer to implicative rules rather than to the more often used Mamdani-Assilian ones. However, it efficiently works if the evaluative linguistic expression with linguistic hedges of the inclusive type are used only, see [36].

The FRBE pre-defined rule base was obtained by an extensive search for association rules on a set of 2829 time series coming from the M3 competition [4], [5].

At a glance, FRBE [30] works as follows:

1) Time series characteristics are computed from the time series being forecasted.
2) The characteristics together with a pre-defined rule base is used to infer the weights of the individual forecasting methods. PbLD is used as the inference method, the inferred outputs are defuzzified using an appropriate defuzzification method DEE, that is again at disposal in the R-package *lfl*.
3) Individual forecasting methods are applied on the time series to obtain individual forecasts.
4) The final forecast is computed as a weighted average of the individual forecasts.

As both of the constructed ensembles combine the above mentioned statistical methods implemented in the R-package *forecast*, we denote them in the subsequent results by R-AVG and R-FRBE, respectively.

III. Participating Methods

A. Adaptive Fuzzy C-Regression Modeling (aFCR)

The *Adaptive Fuzzy C-Regression Modeling* (aFCR) technique [37] is based on the first order Takagi-Sugeno fuzzy model. Each rule of the model consists of a fuzzy set in the antecedent and an affine function in the consequent. Authors construct such Takagi-Sugeno model by learning the rule antecedent using a fuzzy clustering algorithm and by estimating the parameters of the affine function in the consequent using a weighted recursive least squares algorithm.

The clustering algorithm uses prototypes (similar to centroids or cluster centers) to define a cluster. Empirical measures are used to decide whether a new cluster has to be introduced into the model or whether a low quality cluster can be deleted from the model.

The function in the consequent is a weighted average of the inputs whose weights are estimated using the weighted recursive least squares algorithm [38].

TABLE I. THE STRUCTURE OF COMPETITION TIME SERIES

Frequency	Count	Lengths of Time Series	Lengths of Forecasting Horizons
daily	11	212, 518, 909	30 (1×), 90 (8×), 180 (2×)
monthly	45	16 – 105	4 (13×), 6 (15×), 12 (16×), 18 (1×)
quarterly	25	8 – 57, 206	4 (24×), 12 (1×)
yearly	10	18 – 48	4 (2×), 5 (6×), 6 (2×)

The parameters of the model were selected based on simulations, in order to use such parameters that on average provide a highest accuracy among all time series in the competition.

B. A Hybrid Forecast Model Combining Fuzzy Time Series, Linear Regression and a New Smoothing Technique (FTS+S and FTS+LR)

The authors of [39] participated in the competition with two variants of their method. In both cases it is a *Hybrid Forecast Model Combining Fuzzy Time Series and a New Smoothing Technique*. The methods differ in whether the *linear regression* was incorporated or not. We denote the variant with linear regression with acronym FTS+LR and the variant without the regression with FTS+S.

As the technique stems from the fuzzy time series model [21], the numeric values of the time series are first of all sorted in ascending order [39]. Then, the fuzzy c-means algorithm is performed to find clusters of such values. Each original numeric value is then replaced with a label of a cluster to which the value belongs in the highest degree. From that point further, only the cluster labels are processed, i.e., the sequence of numeric values of the original time series is replaced with a sequence of corresponding cluster labels.

The forecast of future values by the FTS+S method is based on last two known values that belong to clusters L_i and L_j, respectively. Forecasted value is computed by finding places in the sequence where the cluster labels L_i and L_j appear consecutively. The cluster labels that appear just after the sequence L_iL_j are then used for prediction by taking the centers of the corresponding clusters and using a smoothing formula to calculate the resulting forecast. This forecasting procedure preserves the main features of the most of the other fuzzy time series procedures, see e.g. [21], [22], which makes it very valuable for an independent comparison.

The FTS+LR method differs in that if no sequence L_i and L_j is found in the time series, then the forecasted value is computed using the linear regression.

C. Multiple Time Series Forecasting based on Fuzzy Techniques (MTSF)

A method entitled *Multiple Time Series Forecasting based on Fuzzy Techniques* (abbr. MTSF) that was introduced in [40] stems from the fuzzy time series approach as introduced by Song [21]. However, there are some crucial differences compared to the original approach [21] or approaches of its many followers [22]–[24]. First of all, time series are classified according to their length and in case of medium or long time series, they are smoothened with the help of the fuzzy transform [8], which helps to extract the trend from the given time series, and consequently, to decompose it into the trend and the seasonal components. Further, the fuzzy time series technique is applied to the smoothened time series. In case of short time series, the technique is applied directly to its values.

Nevertheless, there is not only a single fuzzy time series that is applied either to the original or smoothened time series values, there are actually three approaches applied. The first one is the, say, standard fuzzy time series approach as introduced by Song and Chissom [21], the second one is similar, however, it builds a fuzzy time series of differences of values which may capture the dynamics of given time series with a trend, which is not possible with the original fuzzy time series model. Finally, the third model is based on "fuzzy tendencies", i.e., on linguistic description of the general trend behaviour of a given time series on longer or shorter periods. Fuzzy tendencies actually describe if a given time series on a given time-period is increasing, decreasing, stable, fluctuating or chaotic, and, moreover, what is the strength of such behaviour.

Finally, all three models are competing in an ensemble which does not combine them all together but chooses the most adequate one. This is done again based on the fuzzy tendency of the whole time series, or more precisely, on the similarity of the fuzzy tendency determined on the in-samples and the fuzzy tendency determined on the forecasted out-of-samples. The similar models are called adequate and using the lowest SMAPE values, the ensemble chooses the most adequate model out of the adequate ones.

D. Combination of Fuzzy and Exponential Models (CFESM)

Combination of Fuzzy and Exponential Models (abbr. CFESM) employs a sort of ensemble model, which combines more forecasting models together with the aim of increasing accuracy and lowering risk of choosing an inappropriate model. The particular models used in the ensemble are the fuzzy time series by Song and Chissom [21] with exponential smoothing and models by Viertl [41], while the latter ones are basically statistical models (e.g. again exponential smoothing) with fuzzy sets (fuzzy coefficients, fuzzy data) and fuzzy arithmetic employed.

Therefore, the proposed CFESM method profits from several phenomenons, for example from the forecasting power of standardized statistical methods, either in the original form or in the fuzzified one, or from the typical advantages of ensembling, that were verified on many studies.

IV. RESULTS

The results provided by the competitors were evaluated and SMAPE (1) of their forecasts was computed for each time series. The winning criterion is to have the best average SMAPE computed on all time series. Disregarding benchmark methods, the winner is the CFESM method.

TABLE II. THE RESULTS OF THE CIF COMPETITION – SORTED BY MEAN AND STANDARD DEVIATION OF SMAPE

Method	Mean	Method	Std.Dev.
R-FRBE	0.11808	*R-FRBE*	0.11162
R-AVG	0.12379	*R-AVG*	0.11920
R-ETS	0.12760	*R-ARIMA*	0.11983
R-ARIMA	0.12906	*R-Theta*	0.12323
R-Theta	0.13043	*R-ETS*	0.13221
R-RW	0.14216	*R-RW*	0.13537
CFESM	0.14405	CFESM	0.14620
aFCR	0.15264	MTSF	0.15251
MTSF	0.16621	aFCR	0.15699
FTS+S	0.22930	FTS+S	0.25434
FTS+LR	0.23353	FTS+LR	0.26426

Besides mean SMAPE that indicates a total accuracy of the method, also a standard devidation of SMAPE was evaluated. Standard deviation of SMAPE is an indicator of the robustness or stability of the method. Table II shows the results sorted by SMAPE mean and standard deviation in ascending order (lower values are better).

To assess the results more thoroughly, each pair of methods was examined by the paired Student t test for equality of means. Statistical testing confirmed significant the difference between FTS+S (resp. FTS+LR) and any other method. Moreover, the test proved superiority of the ensembles (R-AVG and R-FRBE) over MTSF and R-RW. Difference among other pairs of method was not statistically significant (at the level of significance $\alpha = 0.05$).

Table III shows the SMAPE means separately for each of daily, monthly, quarterly, and yearly frequency. Although the best methods differ, there was not confirmed any statistically significant difference between them (at 0.05 level of significance). Note that MTSF interestingly won on daily time series.

An interesting view provide results from the forecasting of time series 11 and 16 of the competition. The "ts11" time series is of daily frequency and we have used it to create monthly "ts16" simply by averaging the consecutive windows of 30 values of "ts11". As can be seen in Figure 1, default settings of benchmark statistical methods failed significantly in forecasting "ts11" while the same phenomenon, only captured in monthly frequency, was forecasted quite well by them. Interestingly, the competing method called MTSF provided quite good forecasts on both frequencies. This example shows a potential to further enhance the automatized forecasts to not only take the default frequency but also to try to make the forecasts also on resampled data.

Note also the forecast of "ts11" provided by FTS+S (and FTS+LR). One can immediately see, that the forecast is a copy of the last part of the training data. On one hand, this shows the potential of fuzzy time series to repeated patterns captured on the training data-sets. On the other hand, this may become also a serious drawback as it happened in this case, because forecasts should employ also other phenomenons such as extrapolation. Similarly, both fuzzy time series methods described in [39] failed in the forecast of "ts16" when the decreasing pattern was applied too early although the trend was still increasing.

V. FINDINGS AND CONCLUSIONS

Although the conclusions and findings should not be considered as a general truth because of the limitations of such observations naturally comes from the number of participating submissions, number and specific character of time series chosen for the competition, and finally, from the particular choice of model parameters made by the contestants. On the other hand, we do not have any other independent comparison based on experimental evaluation and thus, these findings are the only ones at disposal at the current time which unquestionably make them valuable. We believe that readers are fully qualified to comprehend the limitations of the findings and to understand the outputs in the given context.

The main findings can be formulated as follows.

1) Standard statistical methods are not that easy to be outperformed as one could guess from many articles on fuzzy approaches.
2) Ensembles give advantages.
3) Fuzzy time series itself in the original setting may face serious problems even with simple time series (e.g. linearly monotone) that go out of the range of the training set.
4) Fuzzy time series may provide some satisfactory forecasting performance, if the computation is different and uses the advantages from other experimentally already sufficiently well confirmed areas such as statistics, ensembling etc.
5) Evaluation in the existing articles possess significant bias for the sake of the presented methods.

Ad1). It is obvious from the results (see Table II), that none of the contestant outperformed standard statistical tools that are at disposal for free in the R-package *forecast* which lowers the current view on fuzzy techniques and their practical applicability. It was not very surprising that it was difficult to beat the standardized and well-tuned statistical tools such as ARIMA or ETS, which are being studied for decades but the fact that RW was not outperformed is not very encouraging. On the other hand, the time series provided by commercial partners (e.g. banks) were very difficult to forecast which made the forecast of sophisticated methods less successful than usually and lowered their results very close to RW. As consequence, none of the methods was statistically significantly better than any other, apart from FTS+LR and FTS+S which were statistically significantly worse than any of the others (benchmarks and contestants). This shows that the victory of statistical benchmarks is not stable enough in order to neglect the fuzzy techniques which should be viewed positively.

Ad2). This conclusion is consistent with findings from the previous studies and competitions including the NN3 [6]. We can only confirm that to the extent of the fact that the two ensembles of statistical benchmarks R-AVG and R-FRBE outperformed all individual benchmarks (although the statistical significance was observed with respect to the R-RW method only). Note that R-FRBE is a method that uses fuzzy rules in order to determine the weights of the ensemble so, this approach aiming at using a fuzzy technique on a higher level not on the forecast itself is an alternative way of how to apply fragments of the fuzzy sets theory in the time series forecasting

TABLE III. MEAN SMAPES PARTITIONED BY FREQUENCIES OF TIME SERIES

Daily	Mean	Monthly	Mean	Quarterly	Mean	Yearly	Mean
MTSF	0.19286	*R-FRBE*	0.08191	*R-Theta*	0.14498	*R-ETS*	0.06982
R-ETS	0.20597	*R-ARIMA*	0.08281	*R-FRBE*	0.14621	*R-AVG*	0.08374
R-FRBE	0.22367	*R-AVG*	0.08631	*R-AVG*	0.15625	*R-RW*	0.08747
CFESM	0.23269	*R-Theta*	0.09454	*R-ETS*	0.17563	*R-FRBE*	0.09438
R-ARIMA	0.23487	*R-ETS*	0.09460	*R-ARIMA*	0.17666	CFESM	0.09812
R-AVG	0.23973	aFCR	0.09803	*R-RW*	0.18092	*R-Theta*	0.10169
R-Theta	0.27034	R-RW	0.10069	CFESM	0.19979	*R-ARIMA*	0.10183
R-RW	0.27344	CFESM	0.10163	aFCR	0.21013	aFCR	0.11553
aFCR	0.27908	MTSF	0.11905	MTSF	0.24989	MTSF	0.13994
FTS+S	0.36438	FTS+S	0.15625	FTS+S	0.31029	FTS+LR	0.19987
FTS+LR	0.36598	FTS+LR	0.16426	FTS+LR	0.31339	FTS+S	0.20697

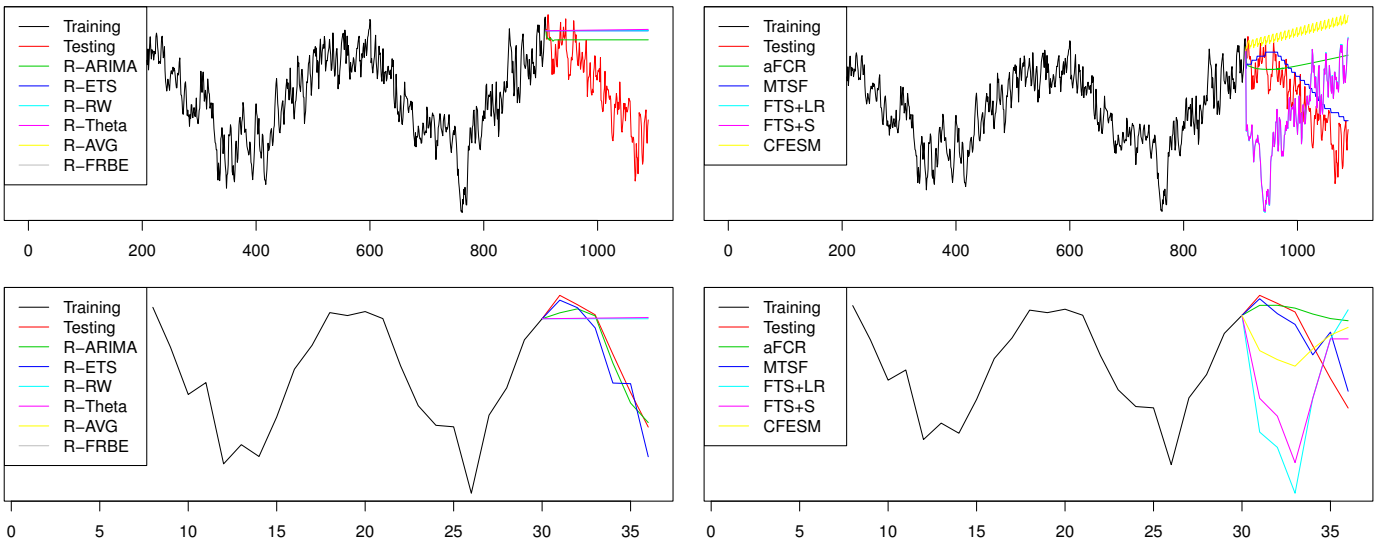

Fig. 1. Results of forecasts on two time series from the competition, "ts11" (top) and "ts16" (bottom), of benchmark methods (left) and competition techniques (right). Time series "ts16", which is displayed in the bottom, is a monthly time series created artificially from the daily time series "ts11" (displayed in the top) by averaging consecutive windows of length 30. While the daily time series was forecasted very poorly even with benchmark methods, the same phenomenon captured with monthly frequency was forecasted quite well. The MTSF is an interesting exception since it captured quite well the correct behaviour of both time series. Note also that the R-FRBE is for "ts11" and "ts16" visible only in case of an extreme zoom-in because its forecasts are very close to those provided by R-AVG and thus, graphically covered by the yellow curve of R-AVG. Similarly, curves of forecasts by FTS+S and FTS+LR coincide in case of "ts11".

task and still preserve the advantages of well-tuned and well-spread statistical methods. Note that the winner – the method denoted by CFESM – also combines more approaches together, namely the fuzzy time series approach [21] and Viertl's statistical approach applied to fuzzy data [41]. The interesting question to be answered is, what would be the forecasting accuracy in case of the use of the Viertl's approach only and the fuzzy time series approach only, to evaluate these two individual methods separately. Also note that although MTSF method, that stems from the fuzzy time series techniques, is a sort of ensemble and was found statistically significantly better than FTS+S and FTS+LR, which confirms the advantages provided by ensembling.

Ad3). Standard fuzzy time series approach as introduced by Song and Chissom [21] and later on elaborated by lots of other authors, obviously possesses several weaknesses. For example, as the original fuzzy time series does describe dependencies between data and does not reflect e.g. differences, it cannot predict even simple time series with a linear increasing or decreasing trend as this trend goes out of the training data range. Therefore, there is lack of potential of this approach to extrapolate trends to out-samples, while either some pattern from the training data-set is repeated (see again the forecasts of "ts11" provided by FTS+S and FTS+LR in Fig. 1) or no rule fires and the forecast is a trivial horizontal line equal. The later phenomenon is confirmed also by the authors in [40], who directly state that "sometimes the forecasting represents the horizontal line, not similar to given time series behaviour".

This weakness has never been observed in vast majority of the previously published articles on fuzzy time series because the experimental evaluation on the training data could not uncover this crucial drawback. Interestingly, the authors of [40] have noticed this drawback and their MTSF, apart from the original fuzzy time series method, employs also fuzzy time series applied on differences of time series values and also fuzzy tendencies, both with the potential to overcome this feedback.

Ad4). As mentioned in the paragraph above, although MTSF employs the fuzzy time series approach, it can overcome the drawback of low extrapolation power and inferring either only observed patterns or horizontal lines, by applying two alternative approaches (differences and fuzzy tendencies)

[40]. The authors also claimed that suggested approach well predicts the time series trends. Although MTSF was not a winner, neither the second contestant, up to some extent this claim can be confirmed, which is visible not only on the presented time series "ts11" and "ts16", but also on the results of daily time series (see Tab. III) where this approach reached the victory and outperformed all the others including benchmarks.

Similarly, the CFESM method uses the fuzzy time series model, but only as an individual method combined with the Viertl's statistical approach with fuzzy data and thus, without a deeper analysis of when which model was chosen in this ensemble and with what success, one cannot state anything about its influence (positive or negative) on the whole CFESM ensemble and its forecasting performance.

It seems that fuzzy time series approaches may compete with the others if being "equipped" with other enriching techniques such as statistical methods or ensembling approaches. But then there are two natural questions to be answered. First, if the fuzzy time series may be competitive ONLY if being enriched? Second, if it makes still sense to call such approaches by the term "fuzzy time series" if the forecasting power comes from different techniques. Neither this paper nor any deeper analysis of the CIF competition can answer these questions, which is the only one more reason to continue in this activity of independent comparisons.

Finally, we owe readers some words on the results obtained by fuzzy regression technique aFCR described in [37]. As well as MTSF or CFESM, this technique did not outperform statistical benchmarks and ensembles. On the other hand, it outperformed MTSF and was not statistically significantly worse than any other. Note that it uses Takagi-Sugeno rules of the first order. If the model based on Takagi-Sugeno rules of a fixed order was (although not better) nearly competitive to the statistical benchmark, we may freely speculate that a rather general approach allowing also other orders and enriched by e.g. ensembling could provide a potential to perform even better.

Of course, we are on the level of speculation. However, the potential of aFCR should not be neglected and some further development of this approach and independent comparison would be definitely of the highest interest.

Acknowledgment

The authors would like to express their gratitude to all the contestants of the CIF competition. Without their courage to enter the independent evaluation, the results, conclusions and findings gathered in this article, would never be obtained. Their sense for fair-play enabled to get information and knowledge which were not at disposal before the competition and which are so desirable. We believe that they will serve as a source of motivation for many followers in consequent competitions.

This work was supported by the European Regional Development Fund in the project of IT4Innovations Centre of Excellence (CZ.1.05/1.1.00/02.0070, VP6).

References

[1] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, "The accuracy of extrapolation (time series) methods: Results of a forecasting competition," *Journal of Forecasting*, vol. 1, pp. 111–153, 1982.

[2] P. Newbold, "The competition to end all competitions," *Journal of Forecasting*, vol. 2, pp. 276–279, 1983.

[3] S. Makridakis, C. Chatfield, M. Hibon, M. Lawrence, T. Mills, K. Ord, and L. F. Simmons, "The m2-competition: A real-time judgmentally based forecasting study," *International Journal of Forecasting*, vol. 9, no. 1, pp. 5 – 22, 1993.

[4] K. Ord, K. Hibon, and S. Makridakis, "The m3–competition," *International Journal of Forecasting*, vol. 16, pp. 433–436, 2000.

[5] S. Makridakis and M. Hibon, "The m3-competition: results, conclusions and implications," *International Journal of Forecasting*, vol. 16, no. 4, pp. 451–476, 2000.

[6] S. F. Crone, M. Hibon, and K. Nikolopoulos, "Advances in forecasting with neural networks? empirical evidence from the {NN3} competition on time series prediction," *International Journal of Forecasting*, vol. 27, no. 3, pp. 635 – 660, 2011.

[7] I. Mrázová and J. Iša, "Growing fuzzy inference neural networks: Their principle and application," in *Intelligent Engineering Systems through Artificial Neural Network*. ASME Press, 2007, pp. 405–410.

[8] I. Perfilieva, "Fuzzy transforms: theory and applications," *Fuzzy Sets and Systems*, vol. 157, pp. 993–1023, 2006.

[9] V. Novák, "Perception-based logical deduction," in *Computational Intelligence, Theory and Applications*, ser. Advances in Soft Computing, B. Reusch, Ed. Berlin: Springer, 2005, pp. 237–250.

[10] V. Novák, M. Štěpnička, A. Dvořák, I. Perfilieva, V. Pavliska, and L. Vavříčková, "Analysis of seasonal time series using fuzzy approach," *International Journal of General Systems*, vol. 37, no. 3, pp. 305–328, 2010.

[11] M. Štěpnička, A. Dvořák, V. Pavliska, and L. Vavříčková, "A linguistic approach to time series modeling with the help of the f-transform," *Fuzzy sets and systems*, vol. 180, pp. 164–184, 2011.

[12] J. Kacprzyk, A. Wilbik, and S. Zadrożny, "Linguistic summarization of time series using a fuzzy quantifier driven aggregation," *Fuzzy Sets and Systems*, vol. 159, no. 12, pp. 1485 – 1499, 2008.

[13] V. Novák, V. Pavliska, M. Štěpnička, and L. Štěpničková, "Time series trend extraction and its linguistic evaluation using f-transform and fuzzy natural logic," in *Recent Developments and New Directions in Soft Computing (Proc. of 2nd World Congress on Soft Computing)*, ser. Studies in Fuzziness and Soft Computing. Springer, 2014, vol. 317, pp. 330–339.

[14] R. Castillo-Ortega, N. Marín, and D. Sánchez, "Time series comparison using linguistic fuzzy techniques," in *Computational Intelligence for Knowledge-Based Systems Design (Proc. of IPMU 2010)*, ser. Lecture Notes in Computer Science, E. Hüllermeier, R. Kruse, and F. Hoffmann, Eds. Berlin, Heidelberg: Springer, 2010, vol. 6178, pp. 330–339.

[15] P. D'Urso and E. Maharaj, "Autocorrelation-based fuzzy clustering of time series," *Fuzzy Sets and Systems*, vol. 160, no. 24, pp. 3565 – 3589, 2009.

[16] J. Aznarte, J. Benítez, and J. Castro, "Smooth transition autoregressive models and fuzzy rule-based systems: Functional equivalence and consequences," *Fuzzy Sets and Systems*, vol. 158, pp. 2734–2745, 2007.

[17] V. Novák, M. Štěpnička, A. Dvořák, I. Perfilieva, V. Pavliska, and L. Vavříčková, "Analysis of seasonal time series using fuzzy approach," *International Journal of General Systems*, vol. 39, pp. 305–328, 2010.

[18] G. Leng, T. McGinnity, and G. Prasad, "An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network," *Fuzzy sets and systems*, vol. 150, pp. 211–243, 2005.

[19] H. J. Rong, N. Sundararajan, G. B. Huang, and P. Saratchandran, "Sequential adaptive fuzzy inference system (safis) for nonlinear system identification and prediction," *Fuzzy Sets and Systems*, vol. 157, pp. 1260–1275, 2006.

[20] N. Kasabov and Q. Song, "Denfis: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction," *IEEE Transactions on Fuzzy Systems*, vol. 10, pp. 144–154, 2002.

[21] Q. Song and B. Chissom, "Fuzzy time series and its models," *Fuzzy Sets and Systems*, vol. 54, pp. 269–277, 1993.

[22] K. Huarng, "Heuristic models of fuzzy time series for forecasting," *Fuzzy Sets and Systems*, vol. 123, pp. 369–386, 2001.

[23] S.-M. Chen and C.-D.Chen, "Taiex forecasting based on fuzzy time series and fuzzy variation groups," *IEEE Transactions on Fuzzy Systems*, vol. 19, pp. 1–12, 2011.

[24] S. Askari and N. Montazerin, "A high-order multi-variable fuzzy time series forecasting algorithm based on fuzzy clustering," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2121 – 2135, 2015.

[25] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1976.

[26] J. D. Hamilton, *Time Series Analysis*. New Jersey: Princeton University Press, 1994.

[27] S. Makridakis, S. Wheelwright, and R. Hyndman, *Forecasting: methods and applications*. USA: John Wiley & Sons, 2008.

[28] R. J. Hyndman, G. Athanasopoulos, S. Razbash, D. Schmidt, Z. Zhou, Y. Khan, and C. Bergmeir, *forecast: Forecasting functions for time series and linear models*, 2013, r package version 4.06. [Online]. Available: http://CRAN.R-project.org/package=forecast

[29] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015. [Online]. Available: http://www.R-project.org/

[30] M. Burda, M. Štěpnička, and L. Štěpničková, "Fuzzy rule-based ensemble for time series prediction: Progresses with associations mining," in *Strengthening Links between Data Analysis and Soft Computing*, ser. Advances in Intelligent Systems and Computing, vol. 315. Heidelberg: Springer, 2014, pp. 261–271.

[31] M. Štěpnička, M. Burda, and L. Štěpničková, "Fuzzy rule base ensemble generated from data by linguistic associations mining," *Fuzzy Sets and Systems*, submitted.

[32] M. Burda, *lfl: Linguistic Fuzzy Logic (R package on CRAN)*, 2015. [Online]. Available: http://cran.r-project.org/web/packages/lfl/

[33] ——, "Linguistic fuzzy logic in R," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2015)*, Istanbul, 2015.

[34] V. Novák and I. Perfilieva, "On the semantics of perception-based fuzzy logic deduction," *International Journal of Intelligent Systems*, vol. 19, pp. 1007–1031, 2004.

[35] A. Dvořák and M. Štěpnička, "On perception-based logical deduction and its variants," in *Proc. 16th World Congress of the International Fuzzy Systems Association and 9th Conference of the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT 2015)*, ser. Advances in Intelligent Systems Research. Gijón: Atlantic Press, 2015.

[36] V. Novák, "A comprehensive theory of trichotomous evaluative linguistic expressions," *Fuzzy Sets and Systems*, vol. 159, no. 22, pp. 2939–2969, 2008.

[37] L. Maciel, A. Lemos, R. Ballini, and F. Gomide, "Adaptive fuzzy c-regressiion modeling for time series forecasting," in *Proc. 16th World Congress of the International Fuzzy Systems Association and 9th Conference of the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT 2015)*, ser. Advances in Intelligent Systems Research. Gijón: Atlantic Press, 2015.

[38] L. Ljung, Ed., *System Identification (2Nd Ed.): Theory for the User*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999.

[39] F. J. J. Santos and H. A. Camargo, "A hybrid forecast model combining fuzzy time series, linear regression and a new smoothing technique," in *Proc. 16th World Congress of the International Fuzzy Systems Association and 9th Conference of the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT 2015)*, ser. Advances in Intelligent Systems Research. Gijón: Atlantic Press, 2015.

[40] T. Afanasieva, N. Yarushkina, M. Toneryan, D. Zavarin, A. Sapunkov, and I. Sibirev, "Time series forecasting using fuzzy techniques," in *Proc. 16th World Congress of the International Fuzzy Systems Association and 9th Conference of the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT 2015)*, ser. Advances in Intelligent Systems Research. Gijón: Atlantic Press, 2015.

[41] R. Viertl, *Statistical methods for fuzzy data*. Wiley, 2011.