# Evaluation of Fuzzy System Ensemble Approach to Predict from a Data Stream

Bogdan Trawiński[1], Magdalena Smętek[1], Tadeusz Lasota[2], and Grzegorz Trawiński[3]

[1] Wrocław University of Technology, Institute of Informatics, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
[2] Wrocław University of Environmental and Life Sciences, Dept. of Spatial Management ul. Norwida 25/27, 50-375 Wrocław, Poland
[3] Wrocław University of Technology, Faculty of Electronics, Wybrzeże S. Wyspiańskiego 27, 50-370 Wrocław, Poland
{bogdan.trawinski,magdalena.smetek}@pwr.wroc.pl, tadeusz.lasota@up.wroc.pl, grzegorz.trawinsky@gmail.com

**Abstract.** In the paper we present extensive experiments to evaluate our recently proposed method applying the ensembles of genetic fuzzy systems to build reliable predictive models from a data stream of real estate transactions. The method relies on building models over the chunks of a data stream determined by a sliding time window and incrementally expanding an ensemble by systematically generated models in the course of time. The aged models are utilized to compose ensembles and their output is updated with trend functions reflecting the changes of prices in the market. The experiments aimed at examining the impact of the number of aged models used to compose an ensemble on the accuracy and the influence of degree of polynomial trend functions applied to modify the results on the performance of single models and ensembles. The analysis of experimental results was made employing statistical approach including nonparametric tests followed by post-hoc procedures devised for multiple $N \times N$ comparisons.

**Keywords:** genetic fuzzy systems, data stream, sliding windows, ensembles, trend functions, property valuation.

## 1 Introduction

Processing data streams poses a considerable challenge because it requires taking into account memory limitations, short processing times, and single scans of arriving data. Many strategies and techniques for mining data streams have been devised. Gaber in his recent overview paper categorizes them into four main groups: two-phase techniques, Hoeffding bound-based, symbolic approximation-based, and granularity-based ones [1]. Much effort is devoted to the issue of concept drift which occurs when data distributions and definitions of target classes change over time [2], [3], [4]. Comprehensive reviews of ensemble based methods for handling concept drift in data streams can be found in [5], [6].

N.T. Nguyen et al. (Eds.): ACIIDS 2014, Part II, LNAI 8398, pp. 137–146, 2014.

For several years we have been investigating methods for generating regression models to assist with real estate appraisal based on fuzzy approach: i.e. genetic fuzzy systems as both single models [7] and ensembles built using various resampling techniques [8], [9]. An especially good performance revealed evolving fuzzy models applied to cadastral data [10], [11]. Evolving fuzzy systems are appropriate for modelling the dynamics of real estate market because they can be systematically updated on demand based on new incoming samples and the data of property sales ordered by the transaction date can be treated as a data stream.

In this paper we present the results of our further study on the method to predict from a data stream of real estate sales transactions based on ensembles of regression models [12], [13], [14], [15]. Having prepared a new real-world dataset we investigated the impact of the number of aged models used to compose an ensemble on the accuracy and the influence of degree of polynomial trend functions applied to modify the results on the performance of single models and ensembles. The scope of extensive experiments was enough to conduct advanced statistical analysis of results obtained including nonparametric tests followed by post-hoc procedures devised for multiple $N\times N$ comparisons.

## 2 Motivation and GFS Ensemble Approach

The approach based on fuzzy logic is especially suitable for property valuation because professional appraisers are forced to use many, very often inconsistent and imprecise sources of information. Their familiarity with a real estate market and the land where properties are located is frequently incomplete. Moreover, they have to consider various price drivers and complex interrelation among them. The appraisers should make on-site inspection to estimate qualitative attributes of a given property as well as its neighbourhood. They have also to assess such subjective factors as location attractiveness and current trend and vogue. So, their estimations are to a great extent subjective and are based on uncertain knowledge, experience, and intuition rather than on objective data. In the paper an evolutionary fuzzy approach to explore data streams to model dynamic real estate market is presented. The problem is not trivial because on the one hand a genetic fuzzy system needs a number of samples to be trained and on the other hand the time window to determine a chunk of training data should be as small as possible to retain the model accuracy at an acceptable level. The processing time in this case is not a decisive issue because property valuation models need not to be updated and/or generated from scratch in an on-line mode.

The outline of the our ensemble approach to predict from a data stream is illustrated in Fig. 1. The data stream is partitioned into data chunks of a constant length $t_c$. The sliding window, which length is a multiple of a data chunk, delineates training sets. We consider a point of time $t_0$ at which the current model was built over data that came in between time $t_0$–$2t_c$ and $t_0$. The models created earlier that have aged gradually are utilized to compose an ensemble so that the current test set is applied to each component model. However, in order to compensate ageing, their output produced for the current test set is updated using trend functions determined over all data since the beginning of the stream. As the functions to model the trends of price changes the polynomials of the degree from one to five were employed: $Ti(t)$, where $i$

denotes the degree. The method of updating the prices of premises with the trends is based on the difference between a price and a trend value in a given time point. More detailed description of the approach presented in the paper can be found in [15].

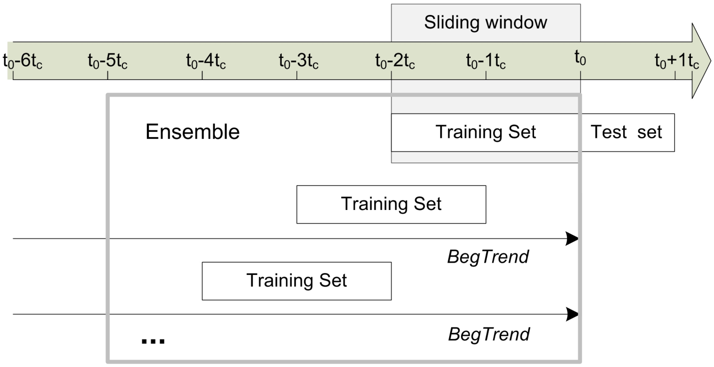


**Fig. 1.** Outline of ensemble approach to predict from a data stream

## 3 Experimental Setup

The investigation was conducted with our experimental system implemented in Matlab devoted to carry out research into ensembles of regression models built over data streams. The data driven models, considered in the paper, were generated using real-world data on sales transactions taken from a cadastral system and a public registry of real estate transactions. Real-world dataset used in experiments was drawn from an unrefined dataset containing above 100 000 records referring to residential premises transactions accomplished in one Polish big city with the population of 640 000 within 14 years from 1998 to 2011. In this period the majority of transactions were made with non-market prices when the council was selling flats to their current tenants on preferential terms. First of all, transactional records referring to residential premises sold at market prices were selected. Then, the dataset was confined to sales transaction data of residential premises (apartments) where the land was leased on terms of perpetual usufruct. The other transactions of premises with the ownership of the land were omitted due to the conviction of professional appraisers stating that the land ownership and lease affect substantially the prices of apartments and therefore they should be used separately for sales comparison valuation methods. The final dataset counted 9795 samples. Due to the fact we possessed the exact date of each transaction we were able to order all instances in the dataset by time, so that it can be regarded as a data stream. Four following attributes were pointed out as main price drivers by professional appraisers: usable area of a flat (*Area*), age of a building construction (*Age*), number of storeys in the building (*Storeys*), the distance of the building from the city centre (*Centre*), in turn, price of premises (*Price*) was the output variable.

The parameters of the architecture of fuzzy systems as well as genetic algorithms are listed in Table 1. Similar designs are described in [7], [16].

**Table 1.** Parameters of GFS used in experiments

| Fuzzy system | Genetic Algorithm |
|---|---|
| Type of fuzzy system: Mamdani | Chromosome: rule base and mf, real-coded |
| No. of input variables: 4 | Population size: 100 |
| Type of membership functions (mf): triangular | Fitness function: MSE |
| No. of input mf: 3 | Selection function: tournament |
| No. of output mf: 5 | Tournament size: 4 |
| No. of rules: 15 | Elite count: 2 |
| AND operator: prod | Crossover fraction: 0.8 |
| Implication operator: prod | Crossover function: two point |
| Aggregation operator: probor | Mutation function: custom |
| Defuzzyfication method: centroid | No. of generations: 100 |

The evaluating experiments were conducted for 37 points of time from 2001-01-01 to 2011-01-01, with the step of 3 months. Component models were built over training data delineated by the sliding windows of constant length of 12 months .The sliding window was shifted by one month along the data stream. The test datasets, current for a given time point, determined by the interval of 3 months were applied to each ensemble. As the accuracy measure the root mean squared error (*RMSE*) was employed. The resulting output of the ensemble for a given time point was computed as the arithmetic mean of the results produced by the component models and corrected by corresponding trend functions.

The analysis of the results was performed using statistical methodology including nonparametric tests followed by post-hoc procedures designed especially for multiple $N \times N$ comparisons [17], [18], [19], [20]. The routine starts with the nonparametric Friedman test, which detect the presence of differences among all algorithms compared. After the null-hypotheses have been rejected the post-hoc procedures should be applied in order to point out the particular pairs of algorithms which produce significant differences. For $N \times N$ comparisons nonparametric Nemenyi's, Holm's, Shaffer's, and Bergamnn-Hommel's procedures are recommended. Due to space limitation the Freidman tests followed by one of the most powerful post-hoc procedures, namely Shaffer's one is consistently applied in the paper.

# 4 Analysis of Experimental Results

## 4.1 Comparison of GFS Ensembles of Different Size

The performance of models of five selected sizes, i.e. single ones and ensembles comprehending 6, 12, 18 and 24 models for no trend correction (*noT*) and *T4* trend functions is illustrated in Figures 2 and 3, respectively. The values of *RMSE* are given in thousand PLN. However, the differences among the models are not visually apparent, therefore one should refer to statistical tests of significance.

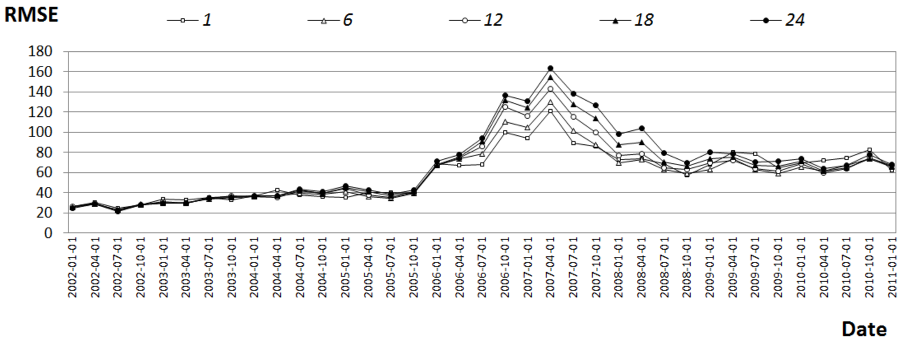


**Fig. 2.** Performance of GFS ensembles of different size for no trend correction (*noT*)

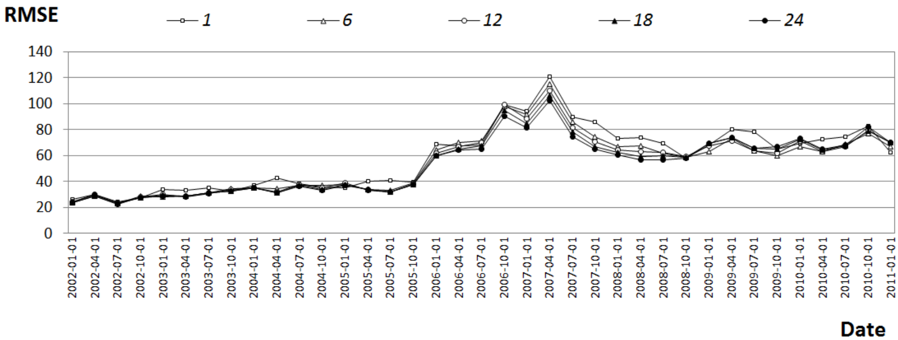


**Fig. 3.** Performance of GFS ensembles of different size for correction with *T4* trend functions

The Friedman test performed in respect of *RMSE* error measure showed that there were significant differences among ensembles in each case. Average ranks of individual models for polynomial trend functions of degrees from *T1* to *T5* produced by the test are shown in Table 2, where the lower rank value the better model. Adjusted *p-values* for the Schaffer's post-hoc procedure are shown in Table 3. The *p-values* indicating statistically significant differences between given pairs of models are marked with italics. The significance level considered for the null hypothesis rejection was assumed 0.05. Following main observations could be done based on the results of both Friedman tests and Shaffer's post-hoc procedures: the greater number of models with corrected outputs with *T4* and *T5* in an ensemble the better performance. The ensembles composed of 15 to 24 models without output correction (*noT*) exhibit the inverse behaviour. The number of null hypotheses rejected by Shaffer's post-hoc procedure out of 36 is equal to 4, 5, 6, 12, 12, and 13 for *T1, T2, T3, T4, T5,* and *noT*, respectively. The following analysis applies only to models with corrected outputs. The models with no output correction reveal statistically worse performance than the ones corrected. The ensembles embracing from 12 to 24 models outperform significantly the single models for *T3, T4*, and *T5* trend functions. Moreover, for *T4* and *T5* trend functions, the ensembles of size from 12 to 24 models surpass single models. In turn, no significant difference is observed among ensembles composed of 6 to 24 models for *T3, T4*, and *T5* trend functions.

**Table 2.** Average rank positions of ensembles of size 1 to 24 produced by Friedman tests

| Trend | 1st | 2nd | 3rd | 4th | 5th | 6th | 7th | 8th | 9th |
|---|---|---|---|---|---|---|---|---|---|
| noT | 6 (3.19) | 3 (3.38) | 9 (3.76) | 12 (4.54) | 1 (4.92) | 15 (5.24) | 18 (6.11) | 21 (6.70) | 24 (7.16) |
| BegT1 | 9 (4.08) | 3 (4.16) | 12 (4.24) | 6 (4.27) | 15 (5.03) | 18 (5.32) | 1 (5.35) | 21 (5.86) | 24 (6.68) |
| BegT2 | 6 (3.78) | 9 (3.84) | 3 (4.27) | 12 (4.51) | 15 (5.03) | 18 (5.38) | 1 (5.57) | 21 (6.14) | 24 (6.49) |
| BegT3 | 24 (4.08) | 21 (4.08) | 18 (4.43) | 15 (4.57) | 6 (4.81) | 12 (5.19) | 9 (5.19) | 3 (5.95) | 1 (7.08) |
| BegT4 | 24 (3.84) | 21 (4.08) | 15 (4.32) | 18 (4.32) | 12 (4.51) | 9 (4.84) | 6 (4.95) | 3 (6.76) | 1 (7.38) |
| BegT5 | 24 (3.81) | 21 (4.00) | 15 (4.30) | 18 (4.35) | 12 (4.54) | 9 (4.95) | 6 (5.00) | 3 (6.68) | 1 (7.38) |

**Table 3.** Adjusted p-values produced by Schaffer's post-hoc procedure for $N\times N$ comparisons for all 36 hypotheses for each degree of trend functions

| Hypotheses | noT | BegT1 | BegT2 | BegT3 | BegT4 | BegT5 |
|---|---|---|---|---|---|---|
| 1 vs 24 | *0.009* | 0.826 | 1.000 | *0.000* | *0.000* | *0.000* |
| 1 vs 21 | 0.107 | 1.000 | 1.000 | *0.000* | *0.000* | *0.000* |
| 1 vs 18 | 0.803 | 1.000 | 1.000 | *0.001* | *0.000* | *0.000* |
| 1 vs 15 | 1.000 | 1.000 | 1.000 | *0.002* | *0.000* | *0.000* |
| 1 vs 12 | 1.000 | 1.000 | 1.000 | *0.010* | *0.000* | *0.000* |
| 1 vs 9 | 0.884 | 1.000 | 0.185 | 0.083 | *0.002* | *0.004* |
| 1 vs 6 | 0.119 | 1.000 | 0.142 | 0.*010* | *0.004* | *0.005* |
| 1 vs 3 | 0.280 | 1.000 | 0.915 | 1.000 | 1.000 | 1.000 |
| 3 vs 24 | *0.000* | *0.002* | *0.014* | 0.095 | *0.000* | *0.000* |
| 3 vs 21 | *0.000* | 0.210 | 0.095 | 0.095 | *0.001* | *0.001* |
| 3 vs 18 | *0.001* | 1.000 | 1.000 | 0.384 | *0.004* | *0.006* |
| 3 vs 15 | 0.075 | 1.000 | 1.000 | 0.669 | *0.004* | *0.005* |
| 3 vs 12 | 0.884 | 1.000 | 1.000 | 1.000 | *0.009* | *0.018* |
| 3 vs 9 | 1.000 | 1.000 | 1.000 | 1.000 | 0.057 | 0.145 |
| 3 vs 6 | 1.000 | 1.000 | 1.000 | 1.000 | 0.098 | 0.187 |
| 6 vs 24 | *0.000* | *0.004* | *0.001* | 1.000 | 1.000 | 1.000 |
| 6 vs 21 | *0.000* | 0.343 | *0.006* | 1.000 | 1.000 | 1.000 |
| 6 vs 18 | *0.000* | 1.000 | 0.270 | 1.000 | 1.000 | 1.000 |
| 6 vs 15 | *0.028* | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 6 vs 12 | 0.507 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 6 vs 9 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 9 vs 24 | *0.000* | *0.002* | *0.001* | 1.000 | 1.000 | 1.000 |
| 9 vs 21 | *0.000* | 0.142 | 0.*009* | 1.000 | 1.000 | 1.000 |
| 9 vs 18 | *0.005* | 1.000 | 0.342 | 1.000 | 1.000 | 1.000 |
| 9 vs 15 | 0.313 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 9 vs 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 12 vs 24 | *0.001* | *0.004* | 0.054 | 1.000 | 1.000 | 1.000 |
| 12 vs 21 | *0.015* | 0.304 | 0.239 | 1.000 | 1.000 | 1.000 |
| 12 vs 18 | 0.249 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 12 vs 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 15 vs 24 | 0.057 | 0.269 | 0.482 | 1.000 | 1.000 | 1.000 |
| 15 vs 21 | 0.350 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 15 vs 18 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 18 vs 24 | 1.000 | 0.744 | 1.000 | 1.000 | 1.000 | 1.000 |
| 18 vs 21 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 21 vs 24 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| **Rejected** | **13** | **4** | **5** | **6** | **12** | **12** |

## 4.2 Comparison of GFS Ensembles Using Trend Functions of Different Degrees

The performance of ensembles comprising models with corrected output using *T1, T2, T3, T4*, and *T5* trend functions and without output correction (*noT*) of sizes from 1 to 24 is illustrated in Figures 4 and 5, respectively. The values of *RMSE* are given in thousand PLN. However, the differences among the models are not visually apparent, therefore one should refer to statistical tests of significance.

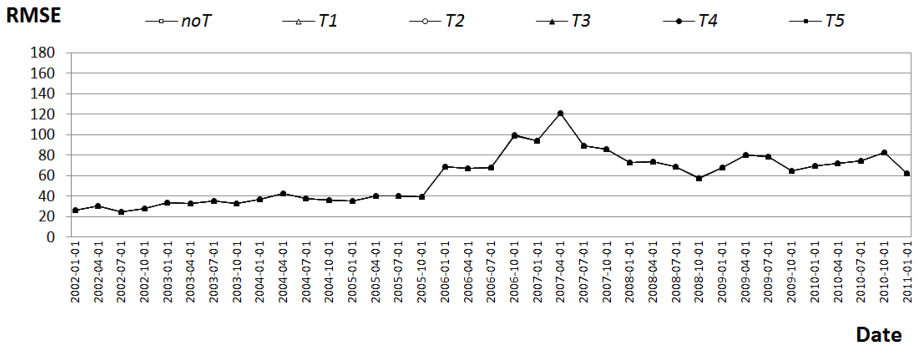


**Fig. 4.** Performance of GFS ensembles with correction using different trend functions for single models

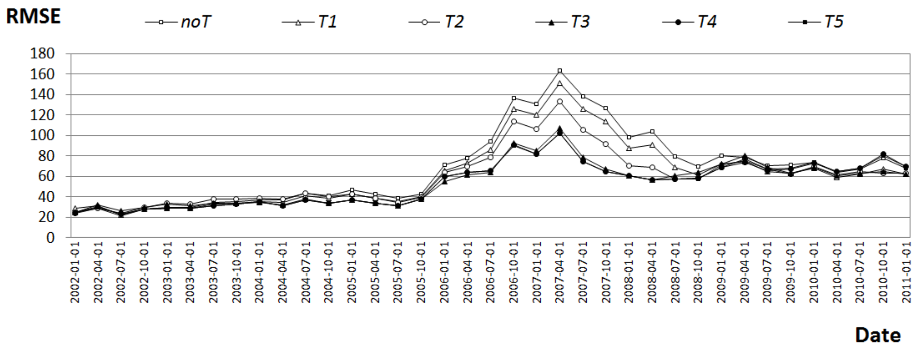


**Fig. 5.** Performance of GFS ensembles with correction using different trend functions for Size=24

The Friedman test performed in respect of *RMSE* values provided by the ensembles showed that there were significant differences among models because *p-value* was much lower than 0.05 in each case. Average ranks of individual models for individual ensemble sizes produced by the test are shown in Table 4, where the lower rank value the better model. Adjusted p-values for one of the Schaffer's post-hoc procedure are shown in Table 5. The *p-values* indicating the statistically significant differences between given pairs of models are marked with italics. Following main observations could be done based on the results of both Friedman tests and Shaffer's post-hoc procedures: the ensembles comprising 3 and more models which output was corrected with *T4* trend functions are in the first position. Next, the ensembles comprising 3, 9 and more models which output was corrected with *T5* trend functions are in the second position, and then the smaller degree of polynomial functions the worse position. However, no significant differences among *T3, T4*, and *T5* trend functions can be observed. The models with no output correction reveal statistically worse performance than the ones corrected with *T3, T4*, and *T5* trend functions for all ensemble sizes. Moreover, there are no significant differences among *T1, T2* trend functions, and *noT*, except for the ensembles composed of 24 models.

**Table 4.** Average rank positions of ensembles with correction using different trend functions determined during Friedman test

| Size | 1st | 2nd | 3rd | 4th | 5th | 6th |
|---|---|---|---|---|---|---|
| 1 | T3 (2.59) | T4 (2.69) | T5 (2.74) | T2 (3.89) | T1 (4.11) | noT (4.97) |
| 3 | T4 (2.57) | T3 (2.73) | T5 (2.78) | T2 (3.95) | T1 (4.05) | noT (4.92) |
| 6 | T4 (2.57) | T5 (2.70) | T3 (2.81) | T2 (3.81) | T1 (4.08) | noT (5.03) |
| 9 | T4 (2.49) | T3 (2.86) | T5 (2.89) | T2 (3.37) | T1 (3.97) | noT (5.05) |
| 12 | T4 (2.46) | T5 (2.86) | T3 (2.97) | T2 (3.78) | T1 (3.86) | noT (5.05) |
| 15 | T4 (2.46) | T5 (2.81) | T3 (2.92) | T2 (3.84) | T1 (3.95) | noT (5.03) |
| 18 | T4 (2.49) | T5 (2.78) | T3 (2.95) | T2 (3.86) | T1 (3.89) | noT (5.03) |
| 21 | T4 (2.51) | T5 (2.78) | T3 (2.97) | T2 (3.84) | T1 (3.84) | noT (5.05) |
| 24 | T4 (2.46) | T5 (2.70) | T3 (3.08) | T2 (3.78) | T1 (3.86) | noT (5.11) |

**Table 5.** Adjusted p-values produced by Schaffer's post-hoc procedure for $N \times N$ comparisons for all 15 hypotheses for each ensemble size

| Hypotheses | sgl | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 24 |
|---|---|---|---|---|---|---|---|---|---|
| noT vs T4 | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* |
| noT vs T5 | *0.000* | *0.000* | *0.000* | *0.000* | *0.001* | *0.000* | *0.000* | *0.000* | *0.000* |
| noT vs T3 | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* |
| T2 vs T4 | *0.040* | *0.015* | *0.035* | *0.043* | 0.242 | *0.015* | *0.015* | *0.023* | *0.012* |
| noT vs T1 | 0.187 | 0.187 | 0.129 | 0.091 | 0.063 | 0.091 | 0.075 | 0.052 | *0.023* |
| T1 vs T4 | *0.011* | *0.006* | *0.005* | *0.006* | 0.128 | *0.006* | *0.012* | *0.023* | *0.023* |
| noT vs T2 | 0.078 | 0.152 | *0.036* | *0.023* | *0.029* | 0.063 | 0.075 | 0.052 | *0.030* |
| T2 vs T5 | 0.058 | 0.053 | 0.076 | 0.281 | 1.000 | 0.128 | 0.091 | 0.108 | 0.053 |
| T1 vs T5 | *0.017* | *0.024* | *0.015* | 0.091 | 1.000 | 0.063 | 0.076 | 0.108 | 0.091 |
| T1 vs T3 | *0.005* | *0.023* | *0.035* | 0.076 | *0.035* | 0.128 | 0.178 | 0.281 | 0.429 |
| T2 vs T3 | *0.020* | *0.036* | 0.129 | 0.281 | 0.063 | 0.139 | 0.178 | 0.281 | 0.429 |
| T3 vs T4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.612 |
| T1 vs T2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| T3 vs T5 | 1.000 | 1.000 | 1.000 | 1.000 | 0.429 | 1.000 | 1.000 | 1.000 | 1.000 |
| T4 vs T5 | 1.000 | 1.000 | 1.000 | 1.000 | 0.856 | 1.000 | 1.000 | 1.000 | 1.000 |
| **Rejected** | **8** | **8** | **8** | **7** | **5** | **5** | **5** | **5** | **7** |

## 5 Conclusions and Future Work

Our further investigation into the method to predict from a data stream of real estate sales transactions based on ensembles of regression models is reported in the paper. The core of our approach is incremental expanding an ensemble by models built from scratch over successive chunks of a data stream determined by a sliding window. In order to compensate ageing the output produced by individual component models for the current test dataset is updated using trend functions which reflect the changes of the market. In our research we employed genetic fuzzy systems of the Mamdani type as the base machine learning algorithms and the trends were modelled over data that came in from the beginning of a stream.

The experiments aimed at examining the impact of the number of aged models used to compose an ensemble on the accuracy and the influence of degree of polynomial trend functions applied to modify the results on the accuracy of single models and ensembles. The data driven models, considered in the paper, were generated using real-world data of sales transactions taken from a cadastral system and a public registry of real estate transactions. The whole dataset, which after cleansing counted 9,795 samples, was ordered by transaction date forming a sort of a data stream. The comparative experiments consisted in generating ensembles of GFS models for 37 points of time within the period of 10 years using the sliding window one year long which delineated training sets. The predictive accuracy of GFS ensembles for different variants of ensemble sizes and polynomial trend functions, was compared using nonparametric tests of statistical significance adequate for multiple comparisons. The ensembles consisted of 3, 6, 9, 12, 15, 18, 21, and 24 component GFS models; for comparison single models were also utilized. As the functions to model the trends of price changes, the polynomials of degree from one to five were employed.

The results proved the usefulness of ensemble approach incorporating the correction of individual component model output. For the majority of cases the bigger ensembles encompassing from 12 to 24 GFS models produced more accurate predictions than the smaller ensembles. Moreover, they outperformed significantly the single models. As for correcting the output of component models, the need to apply trend functions to update the results provided by ageing models is indisputable. However, the selection the most suitable trend function in terms of the polynomial degree has not been definitely resolved. As the matter of fact in majority of cases the trend functions of higher degree, i.e. four and five led to better accuracy. However, the differences were not statistically significant. Therefore, further study is needed for example into the selection of correcting functions dynamically depending on the nature of price changes.

We plan to conduct experiments employing other algorithms capable of learning from concept drifts such as: decision trees, recurrent neural networks, support vector regression, etc. Moreover, we intend to compare the results provided by data-driven regression models with the predictions made by professional appraisers using standard procedures.

**Acknowledgments.** This paper was partially supported by the Polish National Science Centre under grant no. N N516 483840.

## References

1. Gaber, M.M.: Advances in data stream mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2(1), 79–85 (2012)
2. Brzeziński, D., Stefanowski, J.: Accuracy Updated Ensemble for Data Streams with Concept Drift. In: Corchado, E., Kurzyński, M., Woźniak, M. (eds.) HAIS 2011, Part II. LNCS (LNAI), vol. 6679, pp. 155–163. Springer, Heidelberg (2011)
3. Sobolewski, P., Woźniak, M.: Concept Drift Detection and Model Selection with Simulated Recurrence and Ensembles of Statistical Detectors. Journal for Universal Computer Science 19(4), 462–483 (2013)
4. Tsymbal, A.: The problem of concept drift: Definitions and related work. Technical Report. Department of Computer Science, Trinity College, Dublin (2004)

5. Kuncheva, L.I.: Classifier ensembles for changing environments. In: Roli, F., Kittler, J., Windeatt, T. (eds.) MCS 2004. LNCS, vol. 3077, pp. 1–15. Springer, Heidelberg (2004)
6. Minku, L.L., White, A.P., Yao, X.: The Impact of Diversity on Online Ensemble Learning in the Presence of Concept Drift. IEEE Transactions on Knowledge and Data Engineering 22(5), 730–742 (2010)
7. Król, D., Lasota, T., Trawiński, B., Trawiński, K.: Investigation of Evolutionary Optimization Methods of TSK Fuzzy Model for Real Estate Appraisal. International Journal of Hybrid Intelligent Systems 5(3), 111–128 (2008)
8. Kempa, O., Lasota, T., Telec, Z., Trawiński, B.: Investigation of bagging ensembles of genetic neural networks and fuzzy systems for real estate appraisal. In: Nguyen, N.T., Kim, C.-G., Janiak, A. (eds.) ACIIDS 2011, Part II. LNCS (LNAI), vol. 6592, pp. 323–332. Springer, Heidelberg (2011)
9. Lasota, T., Telec, Z., Trawiński, G., Trawiński, B.: Empirical Comparison of Resampling Methods Using Genetic Fuzzy Systems for a Regression Problem. In: Yin, H., Wang, W., Rayward-Smith, V. (eds.) IDEAL 2011. LNCS, vol. 6936, pp. 17–24. Springer, Heidelberg (2011)
10. Lasota, T., Telec, Z., Trawiński, B., Trawiński, K.: Investigation of the eTS Evolving Fuzzy Systems Applied to Real Estate Appraisal. Journal of Multiple-Valued Logic and Soft Computing 17(2-3), 229–253 (2011)
11. Lughofer, E., Trawiński, B., Trawiński, K., Kempa, O., Lasota, T.: On Employing Fuzzy Modeling Algorithms for the Valuation of Residential Premises. Information Sciences 181, 5123–5142 (2011)
12. Trawiński, B., Lasota, T., Smętek, M., Trawiński, G.: An Attempt to Employ Genetic Fuzzy Systems to Predict from a Data Stream of Premises Transactions. In: Hüllermeier, E., Link, S., Fober, T., Seeger, B. (eds.) SUM 2012. LNCS (LNAI), vol. 7520, pp. 127–140. Springer, Heidelberg (2012)
13. Trawiński, B., Lasota, T., Smętek, M., Trawiński, G.: An Analysis of Change Trends by Predicting from a Data Stream Using Genetic Fuzzy Systems. In: Nguyen, N.-T., Hoang, K., Jędrzejowicz, P. (eds.) ICCCI 2012, Part I. LNCS (LNAI), vol. 7653, pp. 220–229. Springer, Heidelberg (2012)
14. Trawiński, B., Lasota, T., Smętek, M., Trawiński, G.: Weighting Component Models by Predicting from Data Streams Using Ensembles of Genetic Fuzzy Systems. In: Larsen, H.L., Martin-Bautista, M.J., Vila, M.A., Andreasen, T., Christiansen, H. (eds.) FQAS 2013. LNCS (LNAI), vol. 8132, pp. 567–578. Springer, Heidelberg (2013)
15. Trawiński, B.: Evolutionary Fuzzy System Ensemble Approach to Model Real Estate Market based on Data Stream Exploration. Journal of Universal Computer Science 19(4), 539–562 (2013)
16. Cordón, O., Herrera, F.: A Two-Stage Evolutionary Process for Designing TSK Fuzzy Rule-Based Systems. IEEE Tr. on Sys., Man and Cyber., Part B 29(6), 703–715 (1999)
17. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30 (2006)
18. García, S., Herrera, F.: An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. Journal of Machine Learning Research 9, 2677–2694 (2008)
19. Graczyk, M., Lasota, T., Telec, Z., Trawiński, B.: Nonparametric Statistical Analysis of Machine Learning Algorithms for Regression Problems. In: Setchi, R., Jordanov, I., Howlett, R.J., Jain, L.C. (eds.) KES 2010, Part I. LNCS (LNAI), vol. 6276, pp. 111–120. Springer, Heidelberg (2010)
20. Trawiński, B., Smętek, M., Telec, Z., Lasota, T.: Nonparametric Statistical Analysis for Multiple Comparison of Machine Learning Regression Algorithms. International Journal of Applied Mathematics and Computer Science 22(4), 867–881 (2012)