Energy 140 (2017) 893–900

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Short-term forecasting of natural gas prices using machine learning and feature selection algorithms

Ervin Čeperić [a], Saša Žiković [b,*], Vladimir Čeperić [c]

[a] HEP Group Inc., Croatia
[b] Faculty of Economics University of Rijeka, Croatia
[c] Faculty of Electrical Engineering University of Zagreb, Croatia

A R T I C L E I N F O

Article history:
Received 16 December 2016
Received in revised form
31 July 2017
Accepted 7 September 2017
Available online 8 September 2017

Keywords:
Natural gas
Henry hub
Machine learning
Feature selection algorithm
Support vector regression machines
Neural networks

A B S T R A C T

We present the results of short-term forecasting of Henry Hub spot natural gas prices based on the performance of classical time series models and machine learning methods, specifically; neural networks (NN) and strategic seasonality-adjusted support vector regression machines(SSA-SVR). We introduce several improvements to the forecasting method based on SVR. A procedure for generation of model inputs and model input selection using feature selection (FS) algorithms is suggested. The use of FS algorithms for automatic selection of model input and the use of advanced global optimization technique PSwarm for the optimization of SVR hyper parameters reduce the subjective inputs. Our results show that the machine learning results reported in the literature often over exaggerate the successfulness of these models since, in some cases, we record only slight improvements over the time series approaches. We have to emphasize that our findings apply to Henry Hub, a market which is known among traders as the “widow maker”. We find definite advantages of using FS algorithms to preselect the variables both in NN and SVR. Machine learning models without the preselection of variables are often inferior to time-series models in forecasting spot prices and in this case FS algorithms show their usefulness and strength.

1. Introduction

Natural gas is the second most actively traded energy commodity, with oil being the number one. With the substitution of coal and wide proliferation of natural gas usage the forecasting of gas prices has become one of the most critical issues in utility' planning and operations. Accurate natural gas price forecasting and the direction of price changes are paramount since these forecasts are used in commodity trading, electricity production planning and regulatory decision making, covering both the supply and demand side of natural gas market. We focus on the short-term forecasting of spot natural gas prices, covering the 1-day ahead and 1-week ahead forecasts.

Due to significant economic effects of price forecasting, many techniques have been investigated [1–3], especially in the electricity related time series (e.g. system load and hourly price forecasts). Some of them are: artificial neural networks (NN) [4–6], Fuzzy Logic Approach [5,7] and many others, including the classical statistical approaches like multiple liner regression and ARMA (autoregressive moving average) [5]. However, publications in the domain of natural gas market oriented forecasts are very scarce. Buchananan et al. [9] is one of the few papers that tries to predict natural gas spot price movements for the US market analysing trader positions which are published on a weekly basis. Nguyen et al. [10] deal with natural gas price forecasts but rely on monthly forward products (futures) instead focusing on spot market prices. By using a multilayer perceptron (MLP), a special form of ANN, they achieved slightly worse results than using the GARCH approach. Another paper analysing natural gas prices is Salehnia et al. [11]. They test several nonlinear models: Local Linear Regression (LLR), Dynamic Local Linear Regression (DLLR) and Artificial Neural Networks (ANN) models. They use daily, weekly and monthly spot prices for Henry Hub from 1997 to 2012. They conclude that MAPE for the DLLR model is lower than for the ANN models, but price prediction of this model is not particularly impressive and quite noisy. Serletis and Shahmoradi [12] analyse the volatility characteristics of Henry Hub gas futures using GARCH modelling. They

* Corresponding author.
E-mail addresses: ervin.ceperic@hep.hr (E. Čeperić), szikovic@efri.hr (S. Žiković), vladimir.ceperic@fer.hr (V. Čeperić).

https://doi.org/10.1016/j.energy.2017.09.026

believe that the large capital requirements and significant lead times, associated with the production and delivery of energy, make these markets very sensitive to the imbalances between demand and supply capability, which results in high price volatility. Moreover, they consider that high volatility is affected by the weather conditions and capacity constraints. Abrishami, Varahrami [13] analyse daily Henry Hub spot price covering 2006–2010 period. They employ a hybrid intelligent framework integrating RES (rule-based expert system) with GMDH neural networks. They claim to obtain better forecasting results compared to the GMDH neural networks and MLF neural networks. Interestingly, the authors claim that the values of their directional change statistic for their hybrid intelligent forecasting method for each evaluation period exceeds 70% (in some periods reaching 92%!). These results are very strange especially since nobody ever came even close to replicating such figures. Busse, Helmholz, Weinmann [14] used a dynamic forecasting approach on basis of a NARX neural network. They investigate the day ahead spot natural gas price in the NetConnect market, Germany. NARX depending on only five factors (temperature, exchange rate and the settlements of three major gas hubs) produces the most accurate forecasts. Although it is the most accurate, compared to the naïve approach NARX approach performed only slightly better. Azadeh et al. [15] tests the performance of Hybrid Neuro-Fuzzy Approach in predicting gas prices in Iran with mixed results. Mishra, Smyth [16] found that natural gas futures prices do not predict the magnitude of future natural gas spot prices any better than what would be predicted by a random walk model. They find that natural gas spot and futures prices are predictable when employing unit root tests which simultaneously allows for heteroskedasticity and multiple structural breaks. On the other hand, findings in the aforementioned literature provide mixed results when applying unit root tests in examining efficient market hypothesis. For example, the existing body of literature claims that the natural gas prices are non-stationary. A number of authors claim that this is a consequence of inadequate treatment of heteroscedasticity in high-frequency data.

During the last decade ANN approaches combined with other methods (usually fuzzy or evolutionary methods) are frequently used in forecasting prices and volumes in the energy markets [2]. On the downside, several problems appear in designing a forecasting system based on ANN for practical purposes. Two problems related to ANN usually appear – first, the “overfitting” (i.e. when an incorrect model fits the data well because its complexity is large enough to the amount of data available) and second, the “curse of dimensionality” (caused by the increase in complexity related to adding extra dimensions to the neural network). Under such conditions, although fitting the past data well, the forecasts can be unsatisfactory. Support vector regression machines (SVR) were proposed by Refs. [17,18] and used by Refs. [19,20] in forecasting to deal with these problems.

Although considered a game changer in the academic community machine learning's performance in real life business application leaves a lot to be desired. Even the tech giants such as Microsoft, Google, and Facebook still consider that the most valuable, high-end machine-learning systems, while useful to them, are still too inflexible and expensive for the companies to offer them to their customers. The problem in application of machine learning systems is that from the purely practical point of view finding patterns is not too hard, but finding the ones that work reliably in the real life situations is. There are several problems with machine learning; if the algorithm is allowed to roam freely through the world of data it can find meaningless patterns which might look robust in the training and validation sets but will fail miserably in the real application. When researchers overcomplicate their models, usually by adding too many parameters to get the results they seek an overfitting problem appears, again leading to spectacular in-sample performance and disappointing out-of-sample results.

The goal of this paper is to add to the gas price forecasting literature in exploring the benefits of machine learning with special focus given to SVR based predications. We also investigate the benefits of using FS preselection of variables to solve the dilemma of using as many variables as possible or focusing only on the most important ones. Feature selection (FS) is the process of selection a subset of relevant features (variables, predictors, inputs) for use in model construction. We test the usefulness of FS both for neural networks and SVR. Furthermore, we develop a generic forecasting strategy with very limited user interaction requirement. We aim to shed light on the question whether SVR, as one of the most advanced forms of machine learning, can significantly contribute to forecasting precision compared to more traditional approaches. To test the successfulness of the machine learning models in short term forecasting of gas prices and add to the literature on machine learning and FS performance we test a very wide range of different combinations of periods, input variables, transformations, window lengths and modelling approaches. We perform the testing of different approaches on the biggest and the most liquid, but also the most unpredictable natural gas market in the world – Henry Hub, US. The volatile nature and the unpredictability of this market is best described by its' colloquial name - the “widow maker”.

The paper is organized in the following manner. In Section 2, the proposed forecasting strategy and modelling method based on SVR is presented. In Section 3 we present the empirical results. Section 4 discussed the obtained results, its implications and concludes.

2. Material and methods

We propose and apply a strategic seasonality-adjusted, support vector machines based model (SSA-SVR) for short-term natural gas price forecasting. Proposed strategy, introduced in Čeperić et al. [21], for short-term natural gas price forecasting consists of:

- SVR machines as the starting point of the model – we choose SVR due to its due favourable properties, discussed in Section 2.1., compared to other functional approximation or regression methods [18,22,23].
- Optimization of SVR hyper parameters by non-derivative based optimization technique, presented in Section 2.1.1. This is done to automate the whole procedure and improve accuracy.
- Generation of model inputs, including seasonal effects and factors influencing the price of natural gas, discussed in Section 2.1.2.
- Selection of model inputs and lags using FS algorithms to additionally automate the model building procedure, presented in Section 2.1.3.

Section 2.1.3 also gives an overview of the proposed strategy for SVR based forecasting is presented, including the description of the steps needed to build a SSA-SVR model.

2.1. Support vector regression

This subsection presents the SVR algorithm. Cortes and Vapnik [24] presented support vector machines (SVR). The idea originated from Vapnik's statistical learning theory [32]. Since then, SVR generated a huge interest in the machine learning literature. The reason for this is their superb performance in solving numerous learning problems, e.g. solving problems in bioinformatics [25] and bankruptcy prediction [26]. Other reasons for the widespread use of SVRs are: a sound theoretical base, low susceptibility to local minima, significant resistance to increased model complexity, usually related to adding dimensions to the model.

SVRs are initially applied for classification problems [24], but soon their application was extended to regression problems [17,18].

The usual formulation of SVR regression is Vapnik's ε-tube support vector regression (SVR) [17,18] and will be briefly introduced, as it is the base of the modelling elaborated in this paper. The first step is to find a linear function *f*, which has the maximum deviation from the training data and simultaneously is as flat as possible:

$$f(\mathbf{x}) = \sum_{i=1}^{l} \boldsymbol{w_i} K(x_i, x) + b \tag{1}$$

where $\mathbf{w}$ is the weight vector, b is a constant and $\mathbf{x}_i$ is the n-dimensional input vector. l is number of training data samples. The flatness of the function f indicates that we seek a small weight vector $\mathbf{w}$. For an overview of the flatness definitions of functions see e.g. Ref. [27]. An easy way to ensure the required flatness is to minimize the following norm $\mathrm{w} = \mathbf{w}, \mathbf{w}$. The support vectors are called the vectors $\mathbf{x}_\mathrm{i}$ corresponding to non-zero $\mathbf{w}_\mathbf{i}$. To calculate weight $\mathbf{w}$, the SVR optimization problem is transferred to the dual optimization problem and solved by allaying quadratic programing [17,18].

All of the stated applies to the linear regression case. To transform the input space into high-dimensional feature space where it is possible to apply the linear SV regression algorithm [22,28] kernel function is used. Kernel function on the two vectors $\mathbf{v}$ and $\mathbf{z}$ is the function $K : X \to \mathbb{R}$ that satisfies:

$$K(\mathbf{v}, \mathbf{z}) = \Phi(\mathbf{v}), \Phi(\mathbf{z}). \tag{2}$$

In this paper radial basis function (RBF) as a kernel function is used:

$$\begin{aligned} K(\mathbf{v}, \mathbf{z}) &= \exp\left(-\gamma \mathbf{v} - \mathbf{z}^2\right) \\ &= exp(-\gamma(\mathbf{v}, \mathbf{v} + \mathbf{z}, \mathbf{z} - 2\mathbf{v}, \mathbf{z})), \text{ for } \gamma > 0. \end{aligned} \tag{3}$$

For the ε-tube SVR the most critical two parameters are:

- The regularization parameter C.
- The width of the ε-tube, i.e. the parameter ε.

The ε-SVR implementation in this paper is based on the LIBSVM tool [29].

2.1.1. Optimization of the SVR hyper-parameters

The selection of SVR hyper parameters is paramount for natural gas price forecasting. E.g. if the value of the hyper parameter C is too big an overfitting phenomenon may appear.

We optimize the SVR parameters on the validation dataset. The validation dataset is chosen to be the closest in time to the test dataset or actual deployment time. Some researchers, e.g. Ref. [30], use cross-validation techniques to optimize model parameters. The logic behind our choice is that if the data is nonstationary using a validation dataset close to the deployment time/testing data set might improve the model accuracy since the model parameters will be optimized on the newest available data. The classical approach to finding a near optimal C and ε is by using the grid search (GS) algorithm [31]. This GS method is used by many researchers in SVR regression, e.g. Ref. [32]. Hsu et al. [33] suggest exponentially growing sequences of SVR parameters as a useful method to find optimal SVR parameters.

The main drawback of the GS approach is the long time needed to perform it [32]. Lately researchers started using the particle swarm optimization method (PSO) to optimize the hyper parameters of the SVR [34–36]. PSO enables fast and accurate identification of near-optimal hyper parameters [34–36] compared to conventional methods such as the GS.

To optimize the SVR hyper parameter we use a PSO inspired technique - particle swarm pattern search method (PSwarm). PSwarm was introduced in Refs. [37,38] and it combines a heuristic approach to global optimization (particle swarm) and pattern search for local minimization. In this way PSwarm has the PSO efficiency and global convergence properties of the pattern search. PSwarm is shown by Refs. [37,38] to be the most robust among a wide range of global optimization solvers and at the same time shows high efficiency.

2.1.2. Data

For the short-term forecasting (1 day and 1 week ahead) of Henry Hub natural gas spot prices we use the following variables: Natural gas prices (Henry Hub NG), Natural gas price differences, Heating oil prices (HO New York harbour), Heating oil price differences, WTI oil prices, WTI oil price differences, Coal prices (COALBSGD - Appalachian mountains), Coal price differences, temperatures in the Northeast US (biggest consumer of NG), Baker Hughes US Natural Gas Rotary Rig Count (BAKG), Total US Natural Gas Marketed Production (NGMPT) (Bcf/d), NG imports from Canada (NAGSCANI) (Bcf/d), effect of days in a week (7 binary variables) and effect of months in a year (12 binary variables). Our training set covers the 2010–2012 period, the validation set covers the 01/01/2013–31/12/2013 period and our testing period covers the 01/01/2014–31/12/2014. All data used in the paper is obtained from Bloomberg.

With regards to the temperature variable three temperature-based sets of variables are used to capture the heating and cooling effects on the price [39]. We use the variable $T_{CR} = \max(T - 20°C, 0)$ to capture the cooling requirement. The heating and additional-heating variables are captured by $T_{HR} = \max(16.5°C - T, 0)$ and $T_{XHR} = \max(5.0°C - T, 0)$. The temperature thresholds represent the standards used by the utility companies [30,39]. For the training purposes we use the past forecasted temperatures instead of the actual measured ones.

An important factor in forecasting natural gas prices is the seasonal variation, e.g. during the winter months' natural gas consumption is higher due to increased heating requirements but also during the hottest summer months due to electricity production from gas-fired power plants. Several options are possible to exploit this factor:

- Use separate model for each season.
- Separate data by using a clustering technique [8].
- Indicating the season by using binary variables [30].

We included additional binary variables in the model - day of week and month in a year to increase the accuracy of our model. We always include these variables since they can implicitly describe the variations in the seasons. Unlike the neural networks the SVR is able to handle the increased dimensionality of the model because the SVR optimization complexity is not bound to the number of model inputs. In our attempts to improve the forecasting results we tried many different combinations of periods, input variables, transformations, window lengths and approaches. Some of the tested modifications:

- Decreasing the number of input characteristics into the FS algorithms as far as to include only the price of gas and its transformations.

- Using logarithms, differences, log differences, square roots, square root differences of variables included in the FS.
- Including the actual and forecasted air temperatures with and without FS algorithm.
- Including the day of the week effects through using binary and single variables.
- Including the month of the year effects through using binary and single variables.
- Performing testing without FS algorithms predominately using lagged values (5 and 10).

We also tried using different testing sets:

- The presented results refer to the four-year period, 2011 and 2012 were used for training the model, 2013 was used for validation and 2014 was used for testing. We tried a combination with just 2012 used as a training set.
- A combination was tried where the validation set (2013) was also used for training the model (three-year training set).
- The whole (four year) set was shortened to three, two and one year.
- Validation set was decreased to 15, 30, 45 and 60 days.
- Models forecasting separately each month in the year were developed and tested – 12 models for one year.
- Models forecasting separately each day in the year were developed and tested using 5, 10, 15, 30 and 60 past days. For validation we used between approximately ½ and ¼ of the data.

Both in case of neural networks and SVR all of these combinations and modifications did not result in any significant improvements and in a lot of cases they provided significantly weaker forecasts. All of these simulations and testing was performed to test as wide a range of options and combinations that can be encountered in machine learning literature. Special focus was given to testing the currently very popular approach in the SVR literature of using very short observation periods for forecasting purposes. This approach also did not yield better forecasting results.

2.1.3. FS algorithms for the selection of model inputs

When building a forecasting model probably the most important step is the selection of model input variables. In practice, the researchers often use their experience to select the variables and that often yields satisfactory results but the whole process requires a lot of time and trial and error attempts. Use of the FS algorithm significantly reduces the required time, automates the modelling process and often improves the model accuracy. We use the FS algorithms as a means to select the near-optimal model inputs without using operator intervention. New inputs are generated by introducing lagged values of the previously mentioned variables.

A number of FS algorithms exists, e.g. FS based on genetic algorithms and information theory (GA&IT) [40], FS using mutual information and maximum-relevance minimum-redundancy criterion (MRMR) [41], FS using Relief algorithm [42] and sequential FS with stepwise regression (Steepwise) [43].

We tested various FS algorithms performance with neural networks and SVR on the validation data set. The validation period is set to begin 1 year before the testing period. The FS algorithms choose 5 or 10 variable values for each analysed variable. The best performer on the validation dataset is the Steepwise FS algorithm. The best FS algorithm on the validation dataset is used to build the model that is tested on the test dataset. The analysis on the test dataset shows that the Steepwise FS is again the best performing FS algorithm, as shown in Tables 1–6.

We use the following steps in building the proposed model:

1. Obtain the data used in the analysis.
2. Generate the model input variables as presented in Tables 1 and 2 Using the FS algorithms select the maximum number of lagged variables. Additional lagged values can be added if the model accuracy is not satisfactory.
3. Split the data between the training and validation sets.
4. Optimize the SVR hyper parameters for each forecast horizon (1 day and 7 days) and number of FS variables (5 and 10) to fit the validation dataset. Each SVR model is built using (only) the training data set. If the model accuracy is unsatisfactory, one can increase the number of lagged values after the FS step and repeat the whole procedure.
5. Test the forecasting model.

3. Results

Basic set of variables consists of NG prices (NG), NG price differences (NG_df), Heating oil prices (HO), Heating oil price differences (HO_df), WTI oil prices (WTI), WTI oil price differences (WTI_df), Coal prices - Appalachian Mountains (COAL), Coal price differences (COAL_df), Baker Hughes US Natural Gas Rotary Rig Count (BAKG), Total US Natural Gas Marketed Production (NGMPT) and NG imports from Canada (Bcf/d) (NAGSCANI). We add days in a week, months in a year effect and temperatures in the NE of US to the Basic set of variables to measure the individual effects of these

Table 1
SVR results with feature selection (5 variables) 1-day ahead forecast.

	Variables	Basic		Days		Months		Temp		All	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
SVR(all)		16,16	2,64	18,78	3,48	23,11	4,02	18,95	3,51	25,17	4,82
SVR SW	NG(1) NG_df(1,8) HO_df(5) NGMPT_df(20)	14,01	2,21	14,18	2,22	14,42	2,29	14,12	2,21	14,10	2,21
SVR MRMR	NG(1,2,4) HO_df(21) BAKG(1)	14,37	2,28	14,28	2,27	14,51	2,32	14,39	2,29	14,37	2,30
SVR GA	NG(1,2) BAKG_df(1,8) COAL(14)	14,05	2,24	14,31	2,23	14,43	2,34	14,18	2,24	14,20	2,26
SVR Relief	NG_df(1-4,6)	15,40	3,92	87,07	16,75	89,51	17,17	79,81	15,50	89,55	17,18
SVR (L5)	NG(1-5)	18,96	2,61	14,25	2,28	14,30	2,27	14,09	2,24	14,19	2,27

[*] Numbers in parenthesis represent the number of lags.
[**] MAPE and RMSE figures are multiplied by 100 for easier viewing.

Table 2
SVR results with feature selection (10 variables) 1-day ahead forecast.

	Variables	Basic		Days		Months		Temp		All	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
SVR (all)		16,16	2,64	18,78	3,48	23,11	4,02	18,95	3,51	25,17	4,82
SVR SW	NG(1) NG_df(3,6) HO(29) HO_df(5,11,14) BAKG_df(29) NGMPT df(17) COAL(30)	14,13	2,25	13,75	2,18	14,01	2,24	13,90	2,26	13,83	2,25
SVR MRMR	NG(1-6,12) NG_df(15) HO(27) BAKG(1)	14,31	2,26	14,42	2,28	14,74	2,36	14,18	2,24	14,36	2,31
SVR GA	NG(1-3,5) BAKG_df(2,3,9-11) COAL(9)	14,26	2,28	14,43	2,31	14,15	2,23	14,15	2,23	14,34	2,29
SVR Relief	NG(1-4) NG_df(1-6)	14,22	2,26	14,17	2,25	14,27	2,24	14,41	2,26	14,37	2,26
SVR(L10)	NG(1-10)	14,76	2,26	14,29	2,24	14,34	2,25	14,59	2,30	14,27	2,26

* Numbers in parenthesis represent the number of lags.
** MAPE and RMSE figures are multiplied by 100 for easier viewing.

Table 3
Comparison of selected models for 1-day ahead forecasts.

	5 variables		10 variables	
	RMSE	MAPE	RMSE	MAPE
SVR(all)	16,16	2,64	16,16	2,64
SVR SW	14,01	2,21	13,75	2,18
NN(all)	17,07	2,93	17,07	2,93
NN SW	14,06	2,41	14,44	2,29
Naïve	14,16	2,25	14,16	2,25
AR(opt)	14,13	2,23	14,13	2,23
ARIMA(opt)	14,10	2,23	14,10	2,23

* MAPE and RMSE figures are multiplied by 100 for easier viewing.

Table 4
SVR results with feature selection (5 variables) 1-week ahead forecast.

	Variables	Basic		Months	
		RMSE	MAPE	RMSE	MAPE
SVR (all)		39,43	6,88	44,06	7,04
SVR SW	NG(7,28) HO(36) BAKG_df(23) COAL(36)	29,04	4,23	34,19	5,43
SVR MRMR	NG(7,9,25) HO_df(18) COAL(7)	38,29	6,67	38,24	6,67
SVR GA	NG(7) BAKG_df(14,15,22) COAL(18)	32,82	5,86	34,48	6,19
SVR Relief	NG_df(9,10,13,14) HO_df(11)	70,14	14,23	w72,63	14,31
SVR(L5)	NG(7-11)	33,95	6,70	33,87	6,69

* Numbers in parenthesis represent the number of lags.
** MAPE and RMSE figures are multiplied by 100 for easier viewing.

Table 5
SVR results with feature selection (10 variables) 1-week ahead forecast.

	Variables	Basic		Months	
		RMSE	MAPE	RMSE	MAPE
SVR (all)		39,43	6,88	44,06	7,04
SVR SW	NG(7,28) HO(22,34) BAKG_df(23,25) NGMPT(23) COAL(27,34)	27,82	4,31	28,85	4,38
SVR MRMR	NG(7-10,13,22) NG_df(24) HO_df(13) NGMPT(30) COAL(13)	32,31	4,89	35,36	5,32
SVR GA	NG(7,10,36) BAKG_df(17,20-25) COAL(16,33)	31,82	4,85	34,86	5,02
SVR Relief	NG_df(7,8,11,13,14) HO(13-17)	75,13	15,04	68,22	14,55
SVR(L10)	NG(7-16)	31,53	4,87	34,86	5,24

* Numbers in parenthesis represent the number of lags.
** MAPE and RMSE figures are multiplied by 100 for easier viewing.

variables. Set "All" consists of all of the aforementioned variables.

The training dataset comprises the data from January 2010 to January 2013. To test the performance of our model, we predict the 1-day ahead and 1-week ahead Henry Hub gas prices in the period January 1st, 2013–December 31st, 2013. We use the data prior to the current forecasting point in our training or validation sets. The validation dataset is fixed at one year prior to the testing period. The selection of the model inputs as well as SVR and NN parameters are optimized on the validation set as described earlier. The proposed model is presented in Tables 1–5. For the error measures of the competing models we use the mean absolute percentage error (MAPE) and root mean square error (RMSE). For easier viewing MAPE and RMSE figures are multiplied by 100. The comparison test data set includes 252 samples (in the 12 month testing period). First we will analyse in detail the performance of different FS algorithms on SVR forecasts (SW – Steepwise, GA - genetic algorithms, MRMR - maximum-relevance minimum-redundancy, Relief - FS using Relief algorithm) and SVR (L5/L10) model based on latest 5/10 lagged values. After obtaining the optimal FS approach for SVR model the same FS approach will be applied to NN and its performance compared to the most data-intensive approach without using FS ("all") and the time series approaches (naïve estimator, autoregressive model with optimal parameters AR(opt), autoregressive integrated moving average with optimal parameters ARIMA(opt)). The results for 1-day ahead forecasts are presented in Tables 1–3

Looking at the results of forecasting 1-day ahead gas prices we

Table 6
Comparison of selected models for 1-week ahead forecasts.

	5 variables		10 variables	
	RMSE	MAPE	RMSE	MAPE
SVR(all)	39,43	6,88	39,43	6,88
SVR SW	29,04	4,23	27,82	4,31
NN(all)	37,15	6,33	37,15	6,33
NN SW	30,22	4,73	28,75	4,55
Naïve	32,36	4,89	32,36	4,89
AR(opt)	31,97	4,80	31,97	4,80
ARIMA(opt)	31,89	4,76	31,89	4,76

* MAPE and RMSE figures are multiplied by 100 for easier viewing.

see a clear dominance of Steepwise (SW) FS algorithm both in terms of RMSE and MAPE statistics. When looking at the five variable based FS's, only in the case of Basic plus Months and Basic plus Temperature variable sets, SVR model based on latest 5 lagged values of natural gas (L5) had a marginally better MAPE statistic. In the case where we used ten variables with the FS algorithms, there is a clear dominance of Steepwise FS across all tested variable sets, both in terms of RMSE and MAPE.

In the five variable FS case there is no improvement, in the sense of RMSE and MAPE statistics, obtained by adding days in a week, months in a year and temperatures (using various combinations – continuous, range based, historical and forecasted). We can conclude that there is no statistically significant gain in adding additional variables to the basic set of variables. In the ten variable FS case there is some gain in accuracy especially in adding days in a week and temperatures variables to the basic set. We can conclude that in the ten variable FS case there is a forecasting gain in considering the effect of days in a week and temperatures in the North East US.

Besides greater or smaller improvements in adding the days and temperatures to the basic set of variables what is much clearer from the results is the superb functioning of FS algorithms. All of the tested FS models scored significantly better results than the SVR/NN model that did not employ a FS preselection of variables. A very significant benefit of using FS algorithms is also visible when running NN models. We have used a feedforward NN trained by Levenberg-Marquardt method and the number of neurons and hidden layers is optimized on the validation data set. What is interesting to see from Table 3 is the marginally better performance of machine learning models with optimized FS compared to traditional time-series models. What is even more surprising is the dominance of classical AR and ARIMA models over NN and SVR models when no preselection of variables was performed. Another interesting find is that the naïve approach (yesterday's value i.e. lag 1) is actually extremely close to the values obtained by AR and ARIMA models and better than both the NN and SVR methods without FS. Machine learning models without the preselection of variables are inferior to traditional time-series models in forecasting day ahead Henry Hub spot prices and in this case the FS algorithms clearly show their usefulness and strength. Out of the analysed variables the most important ones in explaining the 1-day ahead natural gas prices are past natural gas prices, heating oil prices and total US natural gas marketed production. Baker Hughes US natural gas rotary rig count and coal prices (Appalachian Mountains) can also be considered significant in explaining the natural gas prices but to a lesser extent.

We perform the same analysis for 1-week ahead forecasts. The results are presented in Tables 4–6

When forecasting 1-week ahead natural gas prices we see greater difference between the analysed FS algorithms but again the Steepwise (SW) FS clearly dominates over the competition in both the RMSE and MAPE categories. Regarding the AR and ARIMA models we again have a similar situation as in the 1-day ahead forecast, and again the naïve approach (last week's value i.e. lag 1 i.e. 1 week equalling 5 trading days) is very close to the results for the AR and ARIMA models and better than the machine learning methods without FS. Unlike the 1-day ahead forecasts where there was some benefit in adding days in a week, months in a year and temperature effects, in the 1-week ahead forecasts there is no gain in adding months to the basic set of variables, since the addition worsens the RMSE and MAPE statistics. This finding applies to both the five and ten variable FS approach.

For a 1-week ahead case we can conclude that the basic set of variables yields the best results. The value added of using FS algorithms is again clearly visible, same as in the 1-day ahead case with the choice of corrects FS approach being even more important. Machine learning models for a 1-week horizon record much clearer advantage over the classical time series models. When used with FS the SVR model is again the preferred model while being the worst performing model if applied without the preselection of variables. Same as with the SVR model the neural networks also profit from applying FS models. The same as with the 1-day ahead forecast exactly the same variables are significant in explaining the Henry Hub spot natural gas prices: past natural gas prices, heating oil prices, total US natural gas marketed production, Baker Hughes US natural gas rotary rig count and coal prices (Appalachian Mountains).

4. Discussion and conclusion

We investigate and analyse the short-term predictability of Henry Hub spot natural gas prices by using the most widely used approaches to short-term forecasting, including the classical time-series models (naïve, AR, and ARIMA) as well as machine learning (NN and SVR). The improvements we propose are made by taking into account that the successfulness of any method (including the machine learning) to forecast natural gas prices depends on the available data. An additional objective of the proposed approach is to reduce the subjective input in the building of the forecasting model. The proposed use of FS algorithms for automatic model input selection as well as the use of the global optimization PSwarm method to optimize the SVR hyper parameters further reduces the human input. Our primary goal was to investigate whether adding complexity and artificial intelligence significantly improves the forecasting results. We were also looking at the dilemma whether "more is better" in the sense that using more input variables in machine learning models improves the end result. To test the successfulness of the analysed models in the short term forecasting of natural gas prices we test a very wide range of different combinations of periods, input variables, transformations, window lengths and machine learning and FS modelling approaches. We perform our testing on the biggest and the most liquid, but also the most unpredictable natural gas market in the world, also known among the traders as the "widow maker" – Henry Hub, US.

Looking at the results of forecasting 1-day ahead gas prices we find a clear dominance of Steepwise (SW) FS algorithm both in terms of RMSE and MAPE statistics. In the five variable FS case there is no improvement obtained by adding days in a week, months in a year and temperatures (using various combinations – continuous, range based, historical and forecasted). In the ten variable FS case there is a forecasting gain in considering the effect of days in a week and temperatures in the North East US. All of the tested SVR FS models scored significantly better results than the SVR model without using FS approach. A very significant benefit of using FS algorithms is also visible when running NN models, since there is a dominance of time series models over NN and SVR models when no preselection of variables was performed. Machine learning models without the preselection of variables turned out to be inferior to traditional time-series models in forecasting day ahead Henry Hub spot prices and here the FS algorithms clearly show their usefulness and strength. When forecasting 1-week ahead prices we see greater difference in performance between FS algorithms but again the Steepwise FS clearly dominates. Unlike the 1-day ahead forecasts where there was some benefit in adding days, months and temperature effects, in the 1-week ahead forecasts there is no gain in adding months to the basic set of variables and temperatures in this case are completely irrelevant since naturally a temperature from a week ago does not influence tomorrows price. The value added of using FS is again clearly visible, same as in the 1-day ahead case, with the difference that the choice of corrects FS approach is extremely important. When used with FS the SVR model is again preferred while at the same time being the worst performer if applied without FS. Same as with the SVR model the NN also profit from applying FS models. Both in 1-day and 1-week forecasts there is a visible dominance of AR and ARIMA models over NN and SVR models without FS. Surprisingly even the naïve approach is very close to the values obtained by AR and ARIMA models and also better than both the NN and SVR methods without FS. Both for the 1-day and 1-week forecast we find the same variables to be significant in explaining the Henry Hub spot natural gas prices: past natural gas prices, heating oil prices, total US natural gas marketed production, Baker Hughes US natural gas rotary rig count and coal prices (Appalachian Mountains).

For the major part we come to similar conclusions like Busse, Helmholz, Weinmann [14] using NARX in the German gas market. They find that NARX depending on only five factors (temperature, exchange rate and the settlements of three major gas hubs) produces the most accurate forecasts. Although it is the most accurate, compared to the naïve approach NARX approach performed only slightly better. The conclusion of Nguyen et al. [10] and Salehnia et al. [11] are similar in their nature. We report similar findings regarding the fact that in a lot of tested cases the SVR model with just the "basic set" works the best. Even though machine learning (especially SVR approach) shows some improvement over the traditional time-series approaches, in some cases, the benefits are only marginal. Our results run contrary to some of the papers claiming superior performance of NN and SVR based approaches, such as the paper by Abrishami, Varahrami [13].

In our attempts to improve the machine learning forecasting results we tried truly a great number of combinations, including periods, input variables, transformations, window lengths and modelling approaches. Classical logic of defending the obtained results by blaming the omitted variables does not look applicable in this case since adding additional variables does not seem reasonable. We included a truly wide array of potentially important variables (used by both academia and the practitioners) and also the FS approach clearly shows that restricting the number of variables yields significantly better results both for NN and SVR. Our results for the Henry Hub spot prices actually show that, contrary to the popular logic of extensive data-mining, when it comes to short-term forecasting of spot natural gas prices, "less is more".

One promising alley of research could be further research into improving FS algorithms since FS approach has shown to be an extremely valuable and actually the main factor behind improving the forecasting results. FS benefits are twofold since while increasing the precision of forecasts they also reduce the computational time and complexity of the whole process. As for the NN and SVR approaches it seems that they have reached their limits when it comes to short term forecasting of Henry Hub spot prices and it is hard to imagine that any amount of fine tuning could significantly improve the obtained results.

Acknowledgement

Financial support from the Croatian Science Foundation under Grant number IP-2013-11-2203 is gratefully acknowledged.

References

[1] Ying C, Luh PB, Che G, Yige Z, Michel LD, Coolbeth MA, et al. Short-term load forecasting: similar day-based wavelet neural networks. IEEE Trans Power Syst 2010;25(1):322–30.
[2] Feinberg EA, Genethliou D. Load forecasting. In: Chow JH, Wu FF, Momoh J, editors. Applied mathematics for restructured electric power systems. US: Springer; 2005. p. 269–85.
[3] Moser M, Jordan M, Petsche T, editors. Cambridge, MA: MIT Press; 1996. p. 281–7.
[4] Felice MD, Xin Y. Short-term load forecasting with neural network ensembles: a comparative study [application notes]. Comput Intell Mag Springer Sci 2011;6(3):47–56.
[5] Park DC, El-Sharkawi MA, Marks RJ, Atlas LE, Damborg MJ. Electric load forecasting using an artificial neural network. IEEE Trans Power Syst 1991;6(2):442–9.
[6] Liu K, Subbarayan S, Shoults RR, Manry MT, Kwan C, Lewis FI, et al. Comparison of very short-term load forecasting techniques. IEEE Trans Power Syst 1996;19(1):164–73.
[7] Abdel-Aal RE. Short-term hourly load forecasting using abductive networks. IEEE Trans Power Syst 2004;19(1):164–73.
[8] Fan S, Chen C. Short-term load forecasting based on an adaptive hybrid method. IEEE Trans Power Syst 2006;21(1):392–401.
[9] Buchananan W, Hodges P, Theis J. Which way the natural gas price: an attempt to predict the direction of the natural gas spot price movements using trader positions. Energy Econ 2001;23:279–93.
[10] Nguyen HT, Nabney IT. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. Energy 2010;35(9):3674–85.
[11] Salehnia N, Falahib MA, Seifi A, Hossein M, Adelib M. Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis. J Nat Gas Sci Eng 2013;14:238–49.
[12] Serletis A, Shahmoradi A. Returns and volatility in the NYMEX Henry Hub natural gas futures market. OPEC Rev Energy Econ Relat issues 2006:171–86.
[13] Abrishami H, Varahrami V. Different methods for gas price forecasting. Cuad Econ 2011;34:137–44.
[14] Busse S, Helmholz P, Weinmann M. Forecasting day ahead spot price movements of natural gas – an analysis of potential influence factors on basis of a NARX neural network. In: Multikonferenz wirtschaftsinformatik. Braunschweig: Institut für Wirtschaftsinformatik; 2012.
[15] Azadeh A, Sheikhalishahi M, Shahmiri S. A hybrid neuro-fuzzy approach for improvement of natural gas price forecasting in vague and noisy environments: domestic and industrial sectors. In: International Conference on Trends in Industrial and Mechanical Engineering (ICTIME'2012), Dubai; March 2012.
[16] Mishra V, Smyth R. Are natural gas spot and futures prices predictable? Econ Model 2016;54:178–86.
[17] Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support vector regression machines. In: Moser M, Jordan M, Petsche T, editors. Advances in neural information processing systems 9. Cambridge, MA: MIT Press; 1996. p. 155–61.
[18] Vapnik V, Golowich SE, Smola A. Support vector method for function approximation, regression estimation, and signal processing. In: Moser M, Jordan M, Petsche T, editors. Advances in neural information processing systems 9. Cambridge, MA: MIT Press; 1996. p. 281–7.
[19] Mohandes M. Support vector machines for short-term electrical load forecasting. Int J Energy Res 2002;26(4):335–45.
[20] Bo-Juen C, Ming-Wei C, Chih-Jen L. Load forecasting using support vector Machines: a study on EUNITE competition 2001. IEEE Trans Power Syst 2004;19(4):1821–30.
[21] Čeperić E, Čeperić V, Barić A. A strategy for short-term load forecasting by support vector regression machines. IEEE Trans power Syst 2013;28(4):4356–64.
[22] Smola A, Scholkopf B. A tutorial on support vector regression. Stat Comput 2004;14(3):199–222.
[23] Kecman V. Learning and soft computing: support vector machines, neural networks, and fuzzy logic models. Cambridge, MA: MIT Press; 2001.
[24] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.
[25] Ben-Hur A, Ong CS, Sonnenburg S, Schölkopf B, Rätsch G. Support vector machines and kernels for computational biology. PLoS Comput Biol 2008;4(10).
[26] Härdle WK, Moro R, Hoffmann L. Learning machines supporting bankruptcy prediction. Berlin, Germany: Collaborative Research Center 649, Humboldt University; 2010.
[27] Schölkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization, and beyond. Cambridge, MA: MIT Press; 2002.
[28] Boser BE, Guyon IM, Vapnik V. A training algorithm for optimal margin classifiers. In: Proceedings of the 5th annual workshop on computational learning theory (COLT'92). Pittsburgh, PA, US: ACM Press; 1992. p. 144–52.

[29] Chang CC, Lin CJ. A LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011;2(3):1–27.

[30] Espinoza M, Suykens JAK, Belmans R, Moor BD. Electric load forecasting. Control Syst 2007;27(5):43–57.

[31] Vapnik V. The nature of statistical learning theory. New York, NY, US: Springer-Verlag; 1995.

[32] Bao Y, Lu Y, Zhang J. Forecasting stock price by SVMs regression. In: *Artificial intelligence: methodology, systems, and applications*, ser. Lecture notes in computer science. Germany: Springer Berlin/Heidelberg; 2004. p. 295–303.

[33] Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. Department of Computer Science and Information Engineering, National Taiwan University; 2003. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf (Accessed 15 August 2016).

[34] Cao C, Xu J. Short-term traffic flow predication based on PSO-SVM. In: Peng Q, Wang KCP, Qiu Y, Pu Y, Luo X, Shuai B, editors. Proceedings of the first international conference on transportation engineeringvol. 246. Chengdu, China: American Society of Civil Engineers (ASCE); 2007. p. 28.

[35] Jiang MH, Yuan XC. Construction and application of PSO-SVM model for personal credit scoring. In: Proceedings of the 7th international conference on computational science, Part IV: ICCS 2007; 2003. Beijing, China.

[36] Fei S, Liu C, Zeng Q, Miao Y. Application of particle swarm optimization-based support vector machine in fault diagnosis of turbo-generator. In: Proc. 2008 sec. Int. Symp. Intell. Inf. Technol. Applicat. IEEE Comput. Soc. (IITA '08); 2008. p. 1040–4. Washington DC, US.

[37] Vaz AIF, Vicente LN. A particle swarm pattern search method for bound constrained global optimization. J Glob Optim 2007;39(2):197–219.

[38] Vaz AIF, Vicente LN. PSwarm: a hybrid solver for linearly constrained global derivative-free optimization. Optim Methods Softw 2009;24(4–5):669–85.

[39] Engle RF, Granger CWJ, Rice J, Weiss A. Semiparametric estimates of the relation between weather and electricity sales. In: Essays in econometrics. Cambridge, UK: Cambridge University Press; 2001. p. 247–69.

[40] Ludwig O, Nunes U. Novel maximum-margin training algorithms for supervised neural networks. Trans Neur Netw 2010;21(6):972–84.

[41] Brown GA. New perspective for information theoretic feature selection. Mach Learn 2009;5:49–56.

[42] Robnik-Sikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. Mach Learn 2003;53(1–2):23–69.

[43] Hocking RRA. Biometrics invited paper. The analysis and selection of variables in linear regression. Biometrics 1976;32(1):1–49.