Erasmus University Rotterdam

Erasmus School of Economics

Short-term load forecasting using fuzzy regression models based on a multivariate adaptive regression spline

Supervisor

Authors

S.H.L.C.G Vermeulen

B.M. Edelman 478470

Second Assessor

J.A. Oorschot

Bachelor Thesis Econometrics and Operations Research

July 5, 2020

Contents

1 Introduction

One would not be at fault when saying that the systems designed and operated for generating and delivering power to almost 7 billion people around the globe can be considered as an eight world wonder. This complexity, along with the great many factors that influence the system, makes it hard to forecast the load of the network. The forecasts however play an even more crucial role than they do in any other market. This has do to with the peculiar structure of energy demand and supply (Hong 2014)(Xie et al. 2016). While in other industries goods can be stored and supplied at a later stage, electricity has to be supplied at the same moment that it is consumed. This makes it crucial for the forecasts of the load to be as accurate as possible (Khuntia et al. 2016).

To that end a large amount of research has been done to improve the accuracy of the forecasts of the load (Hong 2014) resulting in a field of research with many different models and diverse approaches to the load forecasting (Hong 2010). Meta-analysis are thereby usefull to determine the different positive and negative aspects concerning the different models and to see how they fare compared to each other. One such meta-analysis is done by Nalcaci et al. (2019) who compare three different models: the multivariate adaptive regression spline (MARS) model, an ordinary linear regression (LN) and a Artificial Neural Network (ANN). To extent this research we include a possibilistic linear regression model in this paper to see how it compares to the three methods used by Nalcaci et al. (2019). The conclusion of the paper of Nalcaci et al. (2019) is that the MARS model outperforms the other models both in stability as in accuracy. To make full use of the findings regarding the MARS model and in hoping to improve the performance we will incorporate the MARS model in the possibilistic linear regression. The question we will try to answer is: **What is the effect of fuzzy regression incorporating a multivariate adaptive regression spline on short-term load forecast** In this manner, we will try to improve the analysis of Nalcaci et al. (2019) because it lacks certain forecast models which are proposed in the latest decade and show a lot of promise (Hong 2014) such as the used possibilistic linear regression. The paper will continue with looking at the research that has been done in this area after which the used data will be discussed. The models which will be used will be explained in the methodology and in the result section an overview and in-depth explenation of the results will be given. The paper concludes with discussing those results and giving suggestions to extend the founded results.

2 Literature review

A very extensive overview of several different load forecasting methods as well as a their analysis is given by Hong (2010). The models compared are an ANN model, a linear regression model and a possibilistic linear regression model and different forms of these models. The linear regression model outperforms both models on almost all data sets. Being only outperformed by the possibilistic model a couple of times. It does however indicate that the possibilistic linear regression model has the potential to outperform the ordinary linear regression. Furthermore it mentions the importance of a good fit of the underlying linear model of the fuzzy linear model to make the fuzzy linear model more accurate and stable.

The fuzzy linear model was first proposed by Tanaka et al. (1980) where it was given as an alternative to linear regression. A fuzzy linear model can be used when the underlying assumptions of the linearity of the relationship between the variables can not be verified (Savic & Pedrycz 1991). It depends on fuzzy numbers which are numbers composed of a center and a spread. As error in the model the fuzziness (spread) of the model instead of the error of the observations is used. Consequently the models try to minimize the fuzziness instead of the observation error (Hong 2010).

One paper that does look at the forecasting capabilities of the fuzzy models is Heshmaty & Kandel (1985). He argues in his paper for fuzzy regression models because of the flexibility and the ability to deal with all sorts of data, crisp as well as fuzzy. He makes the case for a good underlying linear model as motor for the fuzzy regression and its potential for forecasting.

A good overview of many fuzzy regression models is given by Chang & Ayyub (2001) where different regression models are shown. Some of them perform exactly the same as the ordinary linear regression in forecasting as for example the fuzzy least squares model minimizing fuzziness. Fuzzy regression minimizing the fuzziness without using the least squares does perform differently from ordinary least squares as also shown by Hong (2010). The paper however does not compare the different models qualitatively.

The potential of fuzzy regression models is also indicated by the paper of Song et al. (2005) where the electrical load is accurately predicted by using a fuzzy linear regression model with an mean average percentage error (MAPE) of 0.036, beating all method proposed by Nalcaci et al. (2019) (MAPE's difference for LN, ANN and MARS +0.011, +0.009 and +0.004 respectively). However, more up-to-date papers indicate that possibilistic linear regression on forecasting leads to disappointing results (Hong & Fan 2016) which have lead to a shortage of literature in the last couple of years.

The overview of Nalcaci et al. (2019) regarding ANN, LN and MARS stipulates the importance of the MARS model in future load forecasting in the comparison to LN and ANN models. The overview however does leave out some of the more advanced methods such as hierarchical forecasting (Hong et al. 2014), echo state networks (Chouikhi et al. 2017) or possibilistic linear models. The inclusion of the the MARS model in the possibilistic framework has not been tried before but shows promise looking at the importance of the underlying linear model in possibilistic models and the superior performance of the MARS model.

3 Data

For the research Dutch load and weather data was used from 2011-1-1 till 2014-12-31. The weather data consisting of the hourly temperature, relative humidity and wind speed and the load data consisting of hourly load for the Netherlands. The weather data is obtained from the Bilt station of the KNMI and the load data is obtained from the ENTSOE. In figure 1 the load is depicted in the first quarter of 2011. The weekends can be clearly seen, as can the repetitive structure of the load. Two patterns of seasonality can be seen in the load, a weekly and a daily pattern. During the week the load is significantly higher than during a weekend. The day-night cycle can also be seen quite clearly with an higher load during daytime than during nighttime, showing in the repetitive pattern of figure 1.

In table 1 the summary statistics are given for the load and the weather data.

Table 1: Summary statistics of the data

	Mean	Standard deviation	Skewness	Kurtosis	Max.	Min.
Load	12713	2338	0.07	1.86	18460	7825
Temperature	10.75	6.54	-0.01	2.97	33.80	-18.80
Humidity	80.45	15.13	-0.85	3.03	100.00	20.00
Windspeed	3.48	1.85	0.80	3.86	14.00	0

We will use 20 different input variables (Nalcaci et al. 2019):

- $x_1, ..., x_{14}$: the lags of the load (hourly, daily, weekly and monthly)(in MW)
- x_{15} national and religious holidays

- x_{16} temperature of the whole day (in a tenth of a Celsius)
- x_{17} relative humidity (in %)
- x_{18} wind speed (in a tenth of a m/s)
- x_{19}, x_{20}: Saturdays and Sundays respectively

The autocorrelation is a slowly declining sine wave with the partial correlations indicating an explanatory power of the first three lags. This is a repeating pattern, the partial correlations of the 24st lag (daily) being significant as well as the 25 and 26 lag which is the same for the second daily lag. Then the partial correlation of the one- and two-week lag and one hour before those are significant as well. Lastly taking into account a monthly lag.

The national and religious holiday variable will be a dummy indicating if an hour falls in a day which has been assigned an official holiday by the Dutch government. The weekend is also captured in the dummy variables x_{19} and x_{20}.

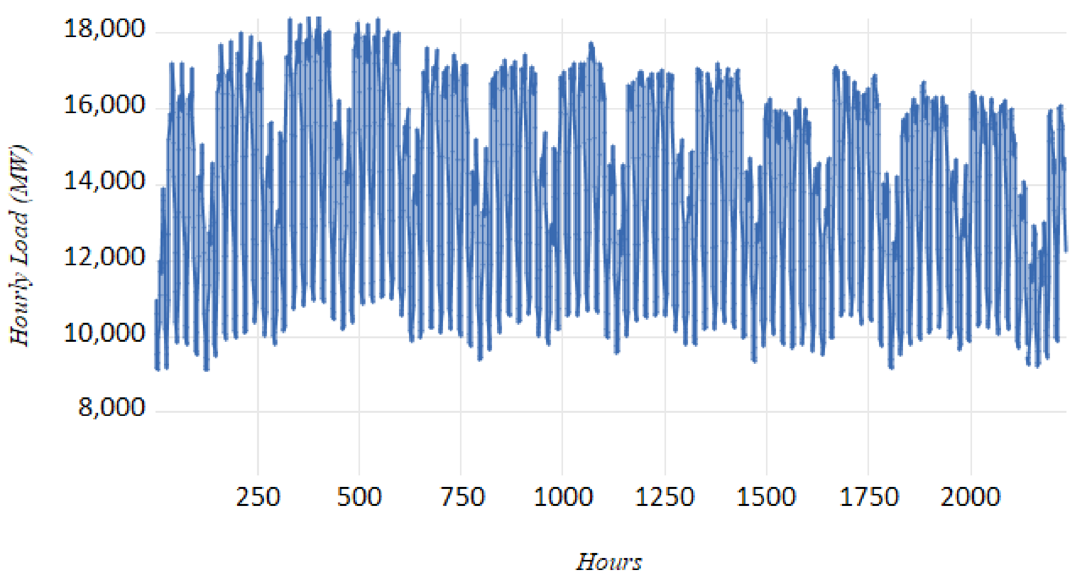


Figure 1: The Dutch hourly load for the first quarter of 2011

4 Theoretical Methodology

Five models will be used in the comparison, the MARS model, the ANN model, a LN model and two fuzzy LN model. We will discuss those all at depth in this section.

4.1 MARS model

A MARS model resembles a linear regression but instead of using the explanatory variables, so-called basic functions are used in the regression (Friedman 1991). Those basis functions can be described accordingly:

$$c^{+}(x,\tau) = max(0, x-\tau), \quad c^{-}(x,\tau) = max(0, \tau - x) \tag{1}$$

where $x, \tau \in \mathbb{R}$. The τ's can be seen as the coordinates of the 'knots' of the complete function where the linearity changes direction or amplitude. This allows for the relationship between the explanatory and dependent variables to be non-linear and include interactive effects. For every explanatory variable $(x_j, (j = 1, 2, ..., p))$ one or more basis function pair (hinge function) will be constructed:

$$\psi := \{c^{+}(x_j,\tau)c^{-}(x_j,\tau) | \tau \in \{x_{1,j}, x_{2,j}, ..., x_{N,j}\} \quad j \in \{1, 2, ..., p\}\} \tag{2}$$

We can then consider the model:

$$Y_i = \theta_0 + \sum_{m=1}^{M} \theta_m T_m(X_i) + \epsilon \tag{3}$$

where M is equal to the amount of basic functions which are used in the model with a maximum of p x N and m indicating the basic function (or combination thereof) which is used. Y is dependent on the basic function T_m as described in (1) and (2) a or combinations of the two of those functions which is given by:

$$T_m(x) = \prod_{j=1}^{K_m} [s_{jm}(x_{j,m} - \tau_{jm})] \tag{4}$$

where s_{jm} is either +1 or -1 , indicating respectively c^{+} or c^{-}. The K_m being the amount of degrees we allow. For example if K_m is equal to one, only basic functions are allowed without interaction effects.

The θ's in (3) can be estimated using ordinary least squares. The most important task is therefore to identify the correct τ's and the hinge functions which will contribute significantly towards the explanatory power of the model. This is done in two steps: a forward and a backward pass.

In the forward pass the hinge functions and corresponding 'knots' are founded. This is done by examining every explanatory variable and every value of that explanatory variable to find a pair of basic functions as described in (2) which causes the largest amount of decrease in the sum of squares of the error terms.

Then the backward pass is used to reduce overfitting in the model culling these hinge functions which contribute the least towards the explanatory power of the network. This is done by using generalized cross-validation (GCV) giving an optimal function f_μ, where the μ indicates the set of founded basic functions in the forward pass. This GCV tries to minimize the following function,:

$$GCV(\mu) = \sum_{i=1}^{N} \frac{(Y_i - f_\mu(x_i))^2}{(1 - P(\mu)/N)^2} \tag{5}$$

where N is the sample amount and $P(\mu)$ the most effective amount of parameters. This formula can than be read as the squared error divided by a penalty depending on the amount of basic functions.

The optimal value of $P(\mu)$ can be calculated using using:

$$P(\mu) = u + d * K \tag{6}$$

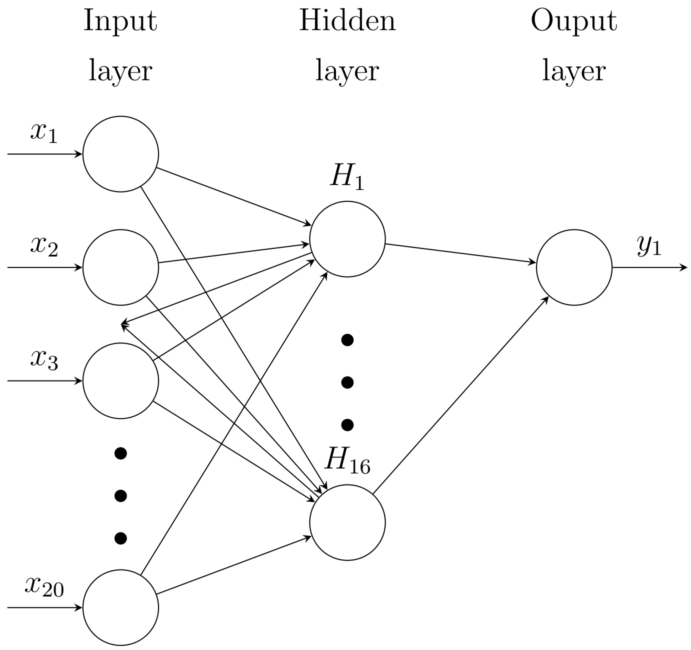
where u is the entire amount of independent hinge functions, K is the number of knots in the model and d being the cost of adding those knots to the model. (Nalcaci et al. 2019).

4.2 Artificial Neural Network

ANN's build on a network of layers filled with neurons and weights connecting the neurons of different layers. In this paper a neural network consisting of an input layer of 20 neurons, an hidden layer of 16 neurons and an output layer of 1 neuron will be used as depicted in figure 1. The weights connecting these neurons 'guide' the initial values through the network using the sigmoid function as activation function of the hidden neurons.

By training the network over the course of 50000 iterations the networks starts to 'recognise' the patterns with which the explanatory variables connect to the dependent variable. This recognition is shown in the way the weights adjust to allow different explanatory variables to have different effects on the dependent variable.

Figure 2: The ANN used



Firstly the weights will be initialized randomly from a uniform(0,1) distribution. Then the weights will be trained using gradient descend using the gradient obtained from the backpropogation method.

4.3 Linear regression

Linear regression is the least complicated and straightforward method which will be used in this paper. The model assumes a linear relationship between the explanatory and dependent variables using the model:

$$Y_i = X_i\beta + \epsilon_i \tag{7}$$

where i = 1,2,...,N. The β will then be estimated using ordinary least squares.

4.4 Possibilistic linear regression

Lastly, possibilistic linear regression will be used as a model (Tanaka et al. 1980)(Hong 2010)(Savic & Pedrycz 1991). This differs from normal linear regression in some aspects as discussed in section 2. Instead of crisp parameters, fuzzy parameters are introduced in (7):

$$Y = A_1x_1 + ... + A_px_p = A\boldsymbol{x} \tag{8}$$

The difference between (7) and (8) being that A is a fuzzy triangular number defined in its membership function as:

$$A_j(a_j) = \begin{cases} 1 - \dfrac{|\alpha_j - a_j|}{c_j} & if \quad \alpha_j = c_j \leq a_j \leq \alpha_j + c_j, \\ 0 & otherwise \end{cases} \tag{9}$$

A fuzzy triangular number as described in (9) consists of a center, the α_j and a spread, the c_j. It can be loosely compared to a probability distribution, consisting of an area rather than a point. A membership function indicating the 'chance' that a certain number is a member of the fuzzy set indicated by a fuzzy number. We can apply this set to crisp number using the extension principle which couples fuzzy numbers to crisp numbers (He et al. 2000) obtaining for every of our crisp numbers y_i a fuzzy set Y_i. This leads to the membership function of y_i, indicating the likeliness of our model's value for y_i to belongs to the fuzzy set Y_i:

$$Y_i(y_i) = \begin{cases} 1 - \dfrac{|y_i - \boldsymbol{x_i}'\boldsymbol{a}|}{\boldsymbol{c}'|\boldsymbol{x_i}|} & for \quad x_i \neq 0, \\ 1 & for \quad x_i = 0, y_i \neq 0, \\ 0 & for \quad x_i = 0, y_i = 0 \end{cases} \tag{10}$$

To obtain the estimates a and c for respectively the center and spread of the variables an linear programming problem is used where the total amount of spread or fuzziness, s, is minimized:

$$s = c_1 + ... + c_m \tag{11}$$

This minimization however does need to make sure to deliver a model which captures the spread of the data to an extend h, which is a chosen threshold. Using (9) and (10) the linear programming problem to find the parameters can be written as (Savic & Pedrycz 1991):

$$\begin{aligned} \text{Minimize} \quad & s = \sum_{j=1}^{m} c_j \\ \text{Subject to} \quad & (1-h)\sum_{j=1}^{m} c_j|x_{ij}| + \boldsymbol{x_i}'\boldsymbol{a} \geq y_i, \\ & (h-1)\sum_{j=1}^{m} c_j|x_{ij}| + \boldsymbol{x_i}'\boldsymbol{a} \geq y_i, \\ \text{for all} \quad & \boldsymbol{c} \geq 0, \quad i = 1, 2, ..., N \end{aligned}$$

In which the parameters $\boldsymbol{a}$ and $\boldsymbol{c}$ are estimated. By including only the spread in the objective function the optimal solution is insensitive to new data points which satisfy the constraints (Savic & Pedrycz 1991).

To make full use of the results of Nalcaci et al. (2019) the possibilistic linear regression is not only done on the explanatory variables but also on the basis functions which are achieved from the MARS model discussed in section 4.1. These basic functions are used to create new explanatory variables $x_{basic,ij}$.The above mentioned linear programming problem then becomes:

$$
\begin{aligned}
\text{Minimize} \quad & s = \sum_{j=1}^{m} c_j \\
\text{Subject to} \quad & (1-h)\sum_{j=1}^{m} c_j x_{basic,ij} + \boldsymbol{x_{basic,i}}'\boldsymbol{a} \geq y_i, \\
& (h-1)\sum_{j=1}^{m} c_j x_{basic,ij} + \boldsymbol{x_{basic,i}}'\boldsymbol{a} \geq y_i, \\
\text{for all} \quad & \boldsymbol{c} \geq 0, \quad i = 1, 2, ..., N
\end{aligned}
$$

5 Empirical methodology

5.1 Evaluation criteria

These explanatory variables will be inserted in the above mentioned models to train them in the period 2011-1-1 till 2012-12-31. The out-of-sample period will be from 2013-11 till 2015-12-31. The performance of the different models is then evaluated using several evaluation criteria given in table 2 for the in-sample as well as the out-of-sample time periods (Nalcaci et al. 2019).

These evaluation criteria give a clear indication of the fitness of the model and the explanatory power. The MAPE, RMSE and MAE indicating the magnitude of the errors while the correlation coefficient and the coefficient of determination indicate how well the hourly load is explained by the models.

5.2 Forecasting

We will discuss two forecasts, an one-hour and an one-day ahead forecast. For the one-hour ahead forecast $\hat{y}_{t+1}$, x_{t+1} will be used. The forecast of the weather is an extensive

Table 2: The evaluation criteria

Abbreviation	Definition
MAE	Mean absolute error
MAPE	Mean absolute percentage error
R^2_{adj}	Coefficient of determination
RMSE	Root mean squared error
r	Correlation coefficient

research area which is beyond the scope of this research. Furthermore the lags and the weekends and holidays are known one hour in advance. Consequently the forecast for the linear and fuzzy linear models will be:

$$\hat{y}_{t+1} = X_{t+1}\beta \tag{12}$$

For the mars model and fuzzy mars model the forecast will be given by:

$$\hat{y}_{t+1} = T(X_{t+1})\theta \tag{13}$$

where (11) and (12) respectively stem from (7) and (3). For the artificial neural network, the trained network from section 4.2 will be used, inputting the explanatory variables x_{t+1} to obtain y_{t+1}

For an one-day forecast a 24 step-ahead forecast will be used. The exogenous variables of the same time ($x_{4:24,t+24}$) will will still be used when computing $\hat{y}_{t+24}$, but for the 1,2 and 3 hour lags the predictions will be used ($\hat{x}_{1:3,t+23}$). In this process we work forward from a prediction for t+1 till a prediction for t+24.

6 Results

6.1 Basic functions

The basic functions which are selected by the MARS-algorithm and used in the model are shown in table 3. It can be seen that the MARS-algorithm uses 5 different explanatory variables, the one-hour, two-hour and three-hour lags, the one-week lag and one-week-and-one-hour lag. The exogenous variables like the different weather conditions, the weekend or public holidays are captured sufficiently by the lags for the model to disregard these altogether. This has as a consequence that the model has to 'wait' on the lags capturing that information before adapting to it, making it slower to react changes in the weather or public holidays.

Furthermore we see that the daily lags are not important for the model, incorporating only the three previous hours and two lags from one week before. This means that in this aspect the model is slow to react, not being able to incorporate a previous day with less hourly load except in the last couple of hours.

However, the model was terminated when adding variables would result in an increase in R^2 lower than 0.0001. This would indicate that the omitted variables do not contribute significantly towards the hourly load.

Table 3: The used basic functions and their coefficients using an one-hour lag (x_1), two-hour lag (x_2), three-hour lag (x_3), one-week lag (x_{10}) and a one-week-and-one-hour lag (x_{11})

Basic functions (coefficients)	
$max(x_{10} - 14530, 0)$(0.840578)	$max(14317 - x_1, 0) * max(x_{11} - 8561, 0)$ (0.000019)
$max(14530 - x_{10}, 0)$(-0.886119)	$max(14317 - x_1, 0) * max(8561 - x_{11}, 0)$ (-0.000193)
$max(x_1 - 14317, 0)$(1.094908)	$max(x_2 - 10840, 0)$(-0.160440)
$max(14317 - x_1, 0)$(-1.083032)	$max(10840 - x_2, 0)$ (-0.201274)
$max(x_{11} - 9723, 0)$(-0.845477)	$max(x_1 - 8486, 0) * max(10840 - x_2, 0)$ (0.000956)
$max(9723 - x_{11}, 0)$(1.136178)	$max(8486 - x_1, 0) * max(10840 - x_2, 0)$ (-0.000443)
$max(14317 - x_1, 0) * max(x_{10} - 8317, 0)$ (-0.000017)	$max(x_2 - 8595, 0) * max(9723 - x_{11}, 0)$ (-0.000078)
$max(14317 - x_1, 0) * max(8317 - x_{10}, 0)$ (0.000174)	$max(8595 - x_2, 0) * max(9723 - x_{11}, 0)$ (0.000006)
$max(x_1 - 8628, 0) * max(9723 - x_{11}, 0)$(0.000265)	$max(x_3 - 9898, 0) * max(x_{11} - 9723, 0)$ (-0.000181)
$max(8628 - x_1, 0) * max(9723 - x_{11}, 0)$ (-0.000831)	$max(9898 - x_3, 0) * max(x_{11} - 9723, 0)$ (0.000224)

6.2 One-hour-ahead forecast

In this section the results regarding the one-hour ahead forecast will be discussed regarding the aforementioned evaluation criteria. In table 4 and figure 3, the in-sample and out-of-sample one-hour-ahead forecasts are shown. The classical methods as discussed in Nalcaci et al. (2019) perform extremely well out-of-sample, explaining more than 99 percent of the variance of the hourly load. This can also be seen by their MAPE, which is lower than 1 percent, and the MAE and RMSE, both giving relatively low values. Interesting is the slightly better performance of the neural network in regards of the MARS-model, contrary to the results of Nalcaci et al. (2019). Both models still outperform the linear model but the neural network seems to deliver a slightly better forecast than the MARS-model.

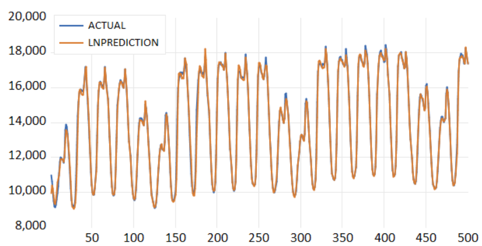
This difference between our results and those of Nalcaci et al. (2019) can be explained by the arbitrariness of neural networks. The amount of iterations during the training are extremely important, inducing underfitting by using to little iterations and overfitting by using to much iterations. Furthermore the initialization of the weights is a determining factor in what minimum the training will terminate, as is the optimization method used or the lay-out of the network. Because of the arbitrary way all those factors play a role in determining the best model, replicating results is highly unlikely.

The fuzzy regression methods perform worse, which was expected for the linear fuzzy regression but goes against expectation for the fuzzy regression based on the MARS-model. As can be seen by their explanatory power which drops to 80 percent and their errors which are an order of magnitude larger than those seen in the other models. Figure 3(d) and figure 3(e) show that the fuzzy models seem to have difficulties capturing the extreme values of the hourly load, underestimating those. Especially noteworthy is the performance of the fuzzy regression based on the MARS-model compared to the normal fuzzy regression. Using the MARS-model leads to a worse performance by the fuzzy regression.

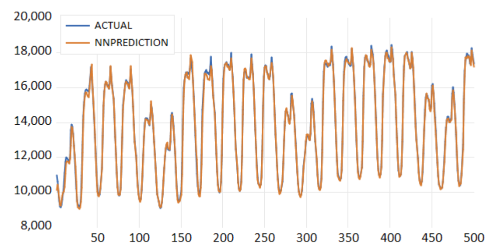
Table 4: The in- and out-of-sample results for a one-hour-ahead forecast with the best result indicated

	Out-of-sample				
	R^2_{adj}	MAPE	MAE	RMSE	r
LinearModel	0.993	0.008	100.59	208.26	0.996
NeuralNetwork	**0.996**	**0.007**	**85.04**	**140.81**	**0.998**
MARSModel	**0.996**	**0.007**	94.36	153.29	**0.998**
FuzzyLinear	0.815	0.063	735.42	1135.94	0.903
FuzzyMars	0.806	0.109	1124.41	2346.66	0.898
	In-sample				
	R^2_{adj}	MAPE	MAE	RMSE	r
LinearModel	0.990	0.008	100.03	227.01	0.995
NeuralNetwork	0.994	**0.006**	**73.50**	**112.96**	0.997
MARSModel	**0.998**	0.008	93.13	155.09	**0.999**
FuzzyLinear	0.822	0.065	734.94	1097.77	0.907
FuzzyMars	0.832	0.144	1420.85	2658.18	0.912

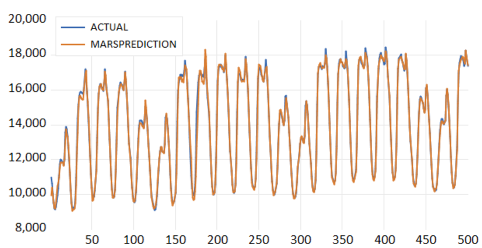
Figure 3: One-hour-ahead forecasts



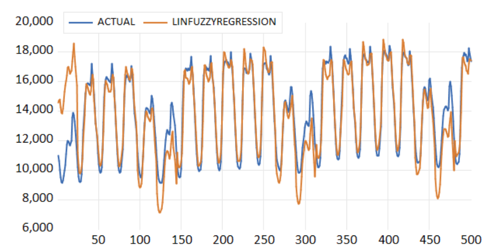
((a)) Linear model



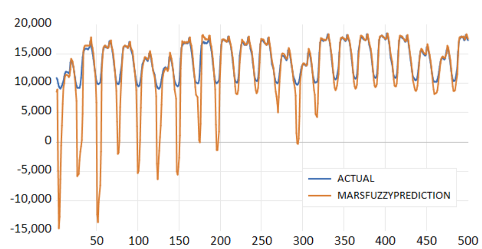
((b)) Neural network



((c)) MARS-model



((d)) Linfuzzy



((e)) MARSfuzzy

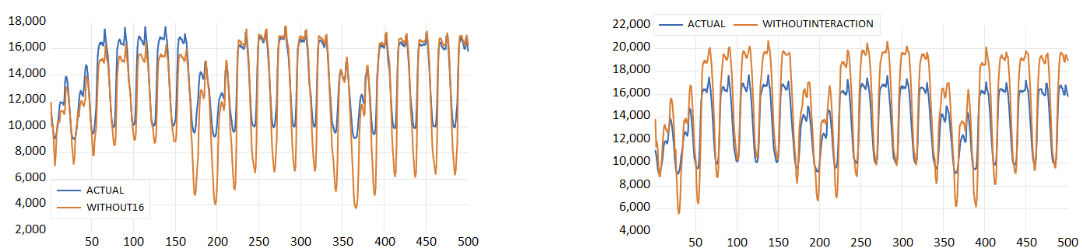
In figure 3(e) one reason for this immediately stands out: the forecast takes on extremely low, sometimes even negative, values in a periodic manner. The forecast is accurate for the high values for the hourly load but drops disproportional when the hourly load drops during the night. This can be explained by use of the basis functions and the coefficients of the model. The basic functions for the interaction effects of the lags have the potential to blow up to more than 20 million because of their inherent multiplication. It seems the fuzzy regression doesn't always have the ability to counter this vast increase in its terms with its coefficients. Especially basic function 16 doesn't seem to be balanced by its coefficient, having a coefficients of -0.0004.

Two possible solutions are the removal of basic function 16 or the removal of interaction effects as a whole. The effect of those two solutions are shown in table 5 and figure 4. It can be seen that the removal of the interaction effects and the removal of the sixteenth basic function does not seem to improve the forecast of the model. The model gives better values for nighttime but during the daytime has become less accurate as depicted by the error terms, averaging 1775 and 2014 respectively.

Table 5: The one-hour-ahead forecast for the fuzzy regression based on the MARS-model without the sixteenth basic function and interaction terms

	R^2_{adj}	MAPE	MAE	RMSE	r
Without16	0.930	0.174	1775.17	2570.21	0.964
WithoutInteraction	0.902	0.170	2014.34	2293.15	0.950

Figure 4: The one-hour-ahead forecast for the fuzzy regression based on the MARS-model without the sixteenth basic function and interaction terms



((a)) Without 16

((b)) Without interaction

The models are almost all robust when looking at the difference between the in-sample and out-of-sample performance in table 4. The results for the three original models and the linear fuzzy regression are almost identical. The only exception is the fuzzy regression based on the MARS-model. The correlation and explained variance increase, however, so do the error terms.

6.3 One-day-ahead forecast

In this section the one-day-ahead forecast will be analyzed. The results are shown in table 6 and figure 5 correspondingly. The forecasts are less accurate than the one-hour-ahead forecast by a large margin. The linear model outperforms all other models, dropping 15 percent in explained variance with regard to the one-hour-ahead forecast and enlarging the errors by a factor five. The other models' explained variances decline much faster.

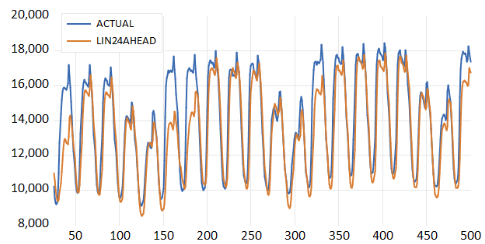
The good performance of the linear model can be explained by the input of the exogenous variables: weekend, holidays, temperature, humidity and wind speed. The MARS-model did not incorporate those variables which makes it heavily reliant on the lags of the hourly load. In a multiple-step-ahead forecast those lags become less reliable because of the prediction error adding up and enlarging the error. This seems to undermine the MARS-model's efficiency in the long run.

Paradoxically, the fuzzy regression based on the MARS model now outperforms the fuzzy regression based on a linear model, indicating that the MARS-model incorporation in the fuzzy regression has some merit. The fuzzy regression seems to place a lower value on the exogenous variables than a normal regression does, relying heavily on the one-hour lag. The MARS fuzzy model seems to spread this reliance more evenly over the different lags, also using one-day and one-week lags.

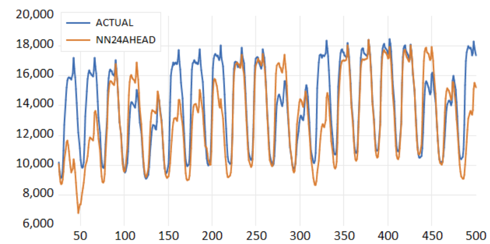
Table 6: The results for an one-day-ahead forecast with the changes in percentages regarding the one-hour-ahead forecast in brackets

	R^2_{adj}	MAPE	MAE	RSME	r
LinearModel	0.843 (-15)	0.051(+538)	655.12 (+551)	991.17(+376)	0.918 (-8)
NeuralNetwork	0.628 (-37)	0.075 (+971)	973.35 (+1045)	1571.43 (+1016)	0.793 (-21)
MARSModel	0.403 (-60)	0.077 (+1000)	986.86 (+946)	2440.69 (+1492)	0.636 (-36)
FuzzyLinear	0.177 (-78)	0.193 (+206)	2387.36 (+225)	3455.62 (+204)	0.422 (-53)
FuzzyMars	0.583 (-28)	0.156 (+46)	1784.60 (+59)	2954.40 (+26)	0.764 (-15)

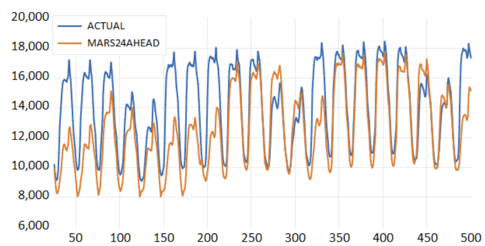
Figure 5: One-day-ahead forecasts



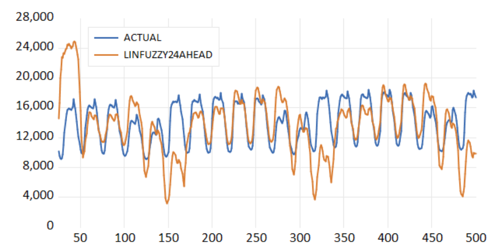
((a)) Linear model



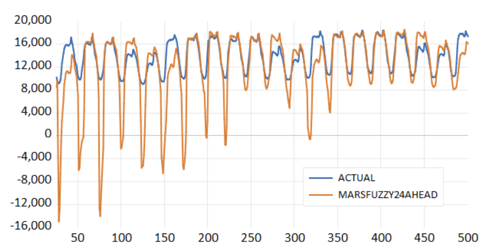
((b)) Neural network



((c)) MARS-model



((d)) Linfuzzy



((e)) MARSfuzzy

7 Discussion

Load-forecasting is an interesting area of research. This can be attributed to the unusual structure of energy demand and supply with no options to store energy. This means forecasts are of extreme importance because the moment energy is generated, it needs to be supplied.

In this paper we compare different forecasting methods for short-term load forecasting. The methods under inspection are a linear regression (LN), an artificial neural network (ANN), a multivariate adaptive regression spline (MARS) and two fuzzy linear regressions, one based on a LN model and one based on a MARS-model. Because of the performance of the MARS model in the paper of Nalcaci et al. (2019) it was expected that incorporating this model in a fuzzy regression would lead to a better performance of the model.

Most models book good results in a one-hour forecast, given error margins lower than one percent and an explanatory power of more than 99 percent. This is, however, not the case for the fuzzy regression models, having an error margin of 6 percent for the fuzzy regression based on a linear model and of 11 percent for the fuzzy regression based on a mars model. The explanatory power of those models dropping to around 80 percent. The models seem to be robust having a comparable in-sample and out-of-sample performance.

A problem for the fuzzy regression based on the MARS model is that it isn't able to limit the huge increase in interaction terms. This has as a consequence that the model suddenly drops heavily to unrealistic values. When removing those interaction terms this problem is solved but the accuracy of the model worsens because of the bad predictions for higher values.

When looking at the one-day ahead forecast we see that the performance of the models fall sharply, in the error terms as well as in the explanatory power. The linear model still performs quite well but the neural network and the MARS-model seems to rely too heavily on the first lag. Especially because of the MARS-model not taking into account the different exogenous variables. The accumulating error of those first lags seem to worsen the performance of those models quite severely. This is also the case for the fuzzy regression based on a linear model while the fuzzy regression based on the MARS-model seems to escape that fate. This can mainly be explained by its more evenly spread reliance on the different lag-terms.

In general we can conclude that fuzzy regression does not seem to be capable of improving short-term load forecasts. The fuzzy regression based on a linear model performing almost a factor ten worse than the three 'classical' models in terms of the errors while dropping twenty percent in explanatory power. The fuzzy regression based on the MARS-model is even further outperformed by all other models, contrary to expectations. This can mainly be explained by the aforementioned unrealistically low values. It would seem fuzzy regression is not able to capture the nonlinear way in which a MARS-model tries to explain the data.

When examining these results one should take into account several factors which could have influenced the accuracy of these results. Firstly, the data used spans only 4 years with 2 years in-sample and 2 years out-of-sample. This is a relatively short time period to build models on, especially with 2011 still being influenced by a financial crisis. Furthermore the weather variables have been measured at a single point in the Netherlands while the weather can differ greatly across the country. These things may make the results less reliable and should caution for unconditional use of the results.

One interesting aspect of the fuzzy regression based on the MARS-model is its ability to be more consistent when doing multiple-steps-ahead forecasts. Its performance in terms of explanatory power even beating that of the ordinary MARS-model. Further research could compare its performance on long-term forecasting.

Furthermore it seems the interaction terms play an interesting role in the fuzzy regression based on the MARS model. On the one hand making it more accurate for higher values but on the other hand making it highly unrealistic for lower values. Further research could look deeper into the effect of the interaction terms on the model, maybe capturing their benefits but losing their drawback.

References

Chang, Y.-H. O. & Ayyub, B. M. (2001), 'Fuzzy regression methods–a comparative assessment', *Fuzzy sets and systems* **119**(2), 187–203.

Chouikhi, N., Ammar, B., Rokbani, N. & Alimi, A. M. (2017), 'Pso-based analysis of echo state network parameters for time series forecasting', *Applied Soft Computing* **55**, 211–225.

Friedman, J. H. (1991), 'Multivariate adaptive regression splines', *The annals of statistics* pp. 1–67.

He, Q., Li, H.-X., Chen, C. P. & Lee, E. (2000), 'Extension principles and fuzzy set categories', *Computers & Mathematics with Applications* **39**(1-2), 45–53.

Heshmaty, B. & Kandel, A. (1985), 'Fuzzy linear regression and its applications to forecasting in uncertain environment', *Fuzzy sets and systems* **15**(2), 159–191.

Hong, T. (2010), 'Short term electric load forecasting'.

Hong, T. (2014), 'Energy forecasting: Past, present and future', *Foresight: The International Journal of Applied Forecasting* **32**, 43–48.

Hong, T. & Fan, S. (2016), 'Probabilistic electric load forecasting: A tutorial review', *International Journal of Forecasting* **32**(3), 914–938.

Hong, T., Pinson, P. & Fan, S. (2014), 'Global energy forecasting competition 2012'.

Khuntia, S. R., Rueda, J. L. & van der Meijden, M. (2016), 'Forecasting the load of electrical power systems in mid- and long-term horizons a review', *IET Generation, Transmission Distribution* **10**(16), 3971–3977.

Nalcaci, G., Özmen, A. & Weber, G. W. (2019), 'Long-term load forecasting: models based on mars, ann and lr methods', *Central European Journal of Operations Research* **27**(4), 1033–1049.

Savic, D. A. & Pedrycz, W. (1991), 'Evaluation of fuzzy linear regression models', *Fuzzy sets and systems* **39**(1), 51–63.

Song, K.-B., Baek, Y.-S., Hong, D. H. & Jang, G. (2005), 'Short-term load forecasting for the holidays using fuzzy linear regression method', *IEEE transactions on power systems* **20**(1), 96–101.

Tanaka, H., Uejima, S. & Asai, K. (1980), 'Fuzzy linear regression model', *International Congress on Applied Systems Research and Cybernetics* .

Xie, J., Chen, Y., Hong, T. & Laing, T. D. (2016), 'Relative humidity for load forecasting models', *IEEE Transactions on Smart Grid* **9**(1), 191–198.