

RESEARCH ARTICLE

Multi-objective regression modeling for natural gas prediction with ridge regression and CMARS

Ayşe Özmen *

Department of Mathematics, TED University, Turkey

Department of Mathematics and Statistics, University of Calgary, Canada

ayseozmen19@gmail.com

ARTICLE INFO

Article history:

Received: 17 February 2021

Accepted: 23 November 2021

Available Online: 1 January 2022

Keywords:

Ridge Regression

(C)MARS

CQP

Interior Point

Prediction of Natural Gas Consumption

AMS Classification 2010:

90C30, 90C51, 90C90

ABSTRACT

Residential customers are the main users generally need a great quantity of natural gas in distribution systems, especially, in the wintry weather season since it is particularly consumed for cooking and space heating. Hence, it ought to be non-interruptible. Since distribution systems have a restricted ability for supply, reasonable planning and prediction through the whole year, especially in winter seasons, have emerged as vital. The Ridge Regression (RR) is formulated mainly to decrease collinearity results through shrinking the regression coefficients and reducing the impact in the model of variables. Conic multivariate adaptive regression splines ((C)MARS) model is constructed as an effective choice for MARS by using inverse problems, statistical learning, and multi-objective optimization theories. In this approach, the model complexity is penalized in the structure of RR and it is constructed a relaxation by utilizing continuous optimization, called *Conic Quadratic Programming (CQP)*. In this study, CMARS and RR are applied to obtain forecasts of residential natural gas demand for local distribution companies (LDCs) that require short-term forecasts, and the model performances are compared by using some criteria. Here, our analysis shows that CMARS models outperform RR models. For one-day-ahead forecasts, CMARS yields a MAPE of about 4.8%, while the same value under RR reaches 8.5%. As the forecast horizon increases, it can be seen that the performance of the methods becomes worse, and for a forecast one week ahead, the MAPE values for CMARS and RR are 9.9% and 18.3%, respectively.



1. Introduction

Local Distribution Companies (LDCs) are the service carriers of private natural gas purchasers within a particular transportation structure. When LDCs target unbalanced gas consumption from Transmission System Operator (TSO) pipelines, regulations, and related policies impose fairly high costs on them in the form of penalties. So correct forecasting is crucial here, as LCD demand is met through spot markets and much of the total demand is organized with global supply contracts through pipelines or liquefied natural gas. The operation must be flexible to compensate for fluctuations in demand. The fluctuation adjustment of energy demand under the constraints of system operation should be realized by predictive models organized for individual types of customers. For LCDs, a unique one-day and one-week forecast gives a discount to cost operations and elimination of penalties

that occur due to unbalanced demand-supply quantities [1,2].

A problem is described as an ill-posed problem if a solution is not unique, present, or stable under perturbation on data. This means that if a small perturbation of the data may bring about a large perturbation of the solution. Ridge Regression is one of the most well-known structures to make these problems regular and stable [3, 4]. It is also known as Tikhonov Regularization. In the statistical literature, there exist some approaches such as principal components regression, partial least squares, least absolute shrinkage and selection operator (LASSO), and ridge regression (RR) to prevent collinearity in traditional linear regression models. Here, RR and LASSO are regularization/penalization strategies that impose a constraint on the regression coefficients while principal components and partial least squares regression are

*Corresponding author

variable subset decision strategies that employ linear combinations of the independent variables in the regression model. RR is formulated mainly to decrease collinearity results by shrinking the regression coefficients and reducing the impact in the model of variables. Here, it seems that shrinking the coefficient estimates may extensively decrease their variance [5].

MARS creates flexible models by employing piecewise linear functions. This continuous model gives a high-quality way to model nonlinearities [6]. Nowadays, as a nonparametric model, MARS is efficiently applied to many areas of technology and science. Here, Conic Multivariate Adaptive Regression Splines (CMARS) model [7-10] is evolved for the backward stage of MARS. It is obtained as a model-based alternative and an effective choice to MARS. In this algorithm, a *Penalized Residual Sum of Squares (PRSS)* is applied for MARS as RR problem and it is worked out by *Conic Quadratic Programming (CQP)*. So, Interior Point Methods [11, 12] and their codes, e.g., MOSEK [13] can be applied by the technique of CMARS. In this paper, we represent natural gas forecasts for one day and one week in advance for a transmission system operator by using RR and CMARS.

In this study, two multi-objective regression models are developed for short-term natural gas demand prediction using RR and CMARS for one day, and one week ahead with daily forecasting intervals. Here, the minimum temperature, the maximum temperature, and heating degree days of the daily average temperature are taken into consideration as different input variables. The MAPE values of CMARS reach 4.8 % and 8.5 for one day, and one week ahead forecasts, respectively. On the other hand, RR gives MAPE values of around 9.9% and 18.3%. We reveal that CMARS performs better than RR in terms of the main performance criteria.

The rest of the paper is organized into four parts. In Sections 2 and 3, we provide a brief review of the models applied in this study. In Section 3, we present our mathematical models. The numerical results of the model are presented and discussed in Section 4. We conclude our study with a discussion of the results and giving future research in Section 5.

2. Ridge regression

The subset determination strategies include applying the least-squares to fit a linear model which involves a subset of the predictors. As an alternative to subset selection, we may fit a model including all k predictors utilizing a procedure that contracts the coefficient estimates towards zero, or equivalently, that compels or regularizes the coefficient estimates [4, 14]. The shrinkage strategies become smaller the regression coefficients by implying a penalty on their size. These techniques bias the estimator of the regression coefficients to decrease the variance in addition to the mean squared error of the estimator and to forestall the model from overfitting [15]. The least-squares fitting procedure estimates $\alpha_0, \alpha_1, \dots, \alpha_k$ using the values that

minimize the Residual Sum of Square (RSS)

$$\sum_{i=1}^n y_i - \alpha_0 - \left(\sum_{j=1}^k \alpha_j x_{ij} \right)^2.$$

The coefficients of RR are the values that minimize the quantity

$$\sum_{i=1}^n \left(y_i - \alpha_0 - \sum_{j=1}^k \alpha_j x_{ij} \right)^2 + \phi \sum_{j=1}^k \alpha_j^2. \quad (1)$$

Here, $\phi \geq 0$ is a *penalty parameter* and it is defined separately. In this method, there is a **tradeoff** between two different criteria (bias and variance). RR search coefficients that match the data well, by decreasing the RSS. Here, the second term, $\phi \sum_{j=1}^k \alpha_j^2$, known as a *shrinkage penalty*, is small if $\alpha_1, \dots, \alpha_k$ are close to zero. Therefore, it affects the estimates of α_j towards zero. The penalty parameter, ϕ , controls the relative effect of these two criteria on the estimation of regression coefficient in (1). When $\phi = 0$, RR produces the least-squares estimates and the penalty term does not affect the model [4, 15]. The penalty term is also known as ℓ_2 penalty and the ℓ_2 norm of a coefficient

vector α is provided by $\|\alpha\|_2 = \sqrt{\sum_{j=1}^k \alpha_j^2}$.

3. Conic multivariate adaptive regression splines

In general, with the CMARS algorithm, implementation of spline function acquires extreme and very important benefits especially in the modeling of dynamics. For both MARS and CMARS, considering a one-dimensional case (input variable), splines are piecewise polynomials. MARS obeys the following general model assumed to exist between the variables [6, 15]:

$$Y = f(\mathbf{x}) + \varepsilon, \quad (2)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_k)^T$ is a vector of predictor, Y is the dependent variable and ε is an error that is supposed to have 0 mean and finite variance. MARS obtains reflected pairs for each input x_j ($j = 1, 2, \dots, k$) with k -dimensional knots $\mathcal{G}_i = (\mathcal{G}_{i,1}, \mathcal{G}_{i,2}, \dots, \mathcal{G}_{i,k})^T$ at each input data vector $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,k})^T$ ($i = 1, 2, \dots, N$). Here, each function is piecewise linear with a knot value, \mathcal{G} . Therefore, in MARS, the Basis Functions (BFs) are determined as [15]

$$C := \left\{ (x_j - \mathcal{G})_+, (\mathcal{G} - x_j)_+ \mid \mathcal{G} \in \{x_{1,j}, x_{2,j}, \dots, x_{N,j}\}, j \in \{1, 2, \dots, k\} \right\}.$$

Here, $f(\mathbf{x})$ in (2) can be closer represented by a successively constructed linear combination of

functions obtained from the set C and Y takes the following form:

$$Y = \alpha_0 + \sum_{m=1}^M \alpha_m \delta_m(\mathbf{x}) + \varepsilon, \quad (3)$$

where α_0 is the intercept, α_m is the unknown coefficient for the m th BF ($m = 1, 2, \dots, M$), and δ_m presents either a function or product of more than one function from the set C . The form of m th BF is represented as follows [15]:

$$\delta_m(\mathbf{x}_i) := \prod_{j=1}^{K_m} (x_{i\kappa_j^m} - \mathcal{G}_{\kappa_j^m})_{\pm} \quad (i = 1, 2, \dots, N). \quad (4)$$

Here, $\mathcal{G}_{\kappa_j^m}$ is the corresponding knot value of the variable $x_{\kappa_j^m}$ and, $x_{i\kappa_j^m}$ is the corresponding input variable for the j th truncated linear function in the m th BF. Moreover, K_m is the interaction degree of the m th BF [16]. To construct the model, the MARS forward stepwise algorithm begins with the constant function $\delta_0(\mathbf{X}^0) = 1$ to estimate α_0 , and all functions in the set C are considered as candidate functions. A lack-of-fit criterion is used to compare the possible BFs that has the form [16,17]:

- 1,
- x_a ,
- $[x_a - \tau_j]_+$,
- $x_a x_b$, and
- $[x_a - \tau_j]_+ [x_b - \tau_k]_+$.

The BFs above employ different input variables, x_a and x_b , with their knots, τ_j and τ_k as input variables cannot be the same for each BF in the MARS algorithm. At each step, with one of the reflected pair in set C , all products of a function $\delta_m(\mathbf{X}^m)$ in the model set are taken into account as a new function pair and added the term to the model set. This term has the following form

$$\alpha_{M+1} \delta_a(\mathbf{X}^a) \cdot [X_k - \tau]_+ + \alpha_{M+2} \delta_a(\mathbf{X}^a) \cdot [\tau - X_k]_+$$

where α_{M+1} and α_{M+2} are coefficients that are estimated by least-squares with all other $M+1$ coefficients. Then the “winning” products are added into the model and then, for example, the possible candidates for BFs as follows [17]:

- 1,
- x_a ,
- $[x_a - \tau_j]_+$, if x_a is in the model already,
- $x_a x_b$, if x_a and x_b are in the model already,
- $[x_a - \tau_j]_+ [x_b - \tau_k]_+$, if $[x_a - \tau_j]_+ x_b$ and

$[x_b - \tau_k]_+ x_a$ are in the model already.

The model obtained by forward stage overfits the data. Therefore, a backwards-pruning procedure is applied to find BFs that contribute least to model fit, and then these BFs are progressively deleted. This iterative process continues until an optimal number of terms are presented in the final model [18, 19]. Here, MARS used a modified recursive partitioning strategy to simplify high-dimensional problems. The sequence of models generated from this procedure is evaluated by using generalized cross-validation (GCV), and the model with the best predictive fit is selected. Consequently, an estimated best model \hat{f}_β of each number of terms, β , is found at the end of that process. In MARS model, generalized cross-validation is applied to define the optimal number of terms, μ , and it also shows the lack of fit. The GCV criterion defined by Friedman [4] is given by

$$LOF(\hat{f}_\beta) = GCV(\beta) := \frac{\sum_{i=1}^N (y_i - \hat{f}_\beta(\mathbf{x}_i))^2}{(1 - M(\beta) / N)^2}.$$

Here, N is the number of sample observations, and $M(\beta)$ represents the effective number of parameters in the model.

In the CMARS method, firstly, the large model provided by the forward MARS algorithm is built up and addressed. Instead of the backward stage algorithm of MARS, the $PRSS$ with M_{\max} BFs is employed as a refinement of the least-squares estimation (LSE) to control the lack of fit from the viewpoint of the tradeoff between goals of *complexity* and *stability* to estimate and assess the function $f(\mathbf{x})$ in (2). Consequently, $PRSS$ can be summed up during the forward stage of MARS and written as [6]:

$$PRSS := \sum_{i=1}^N (y_i - \boldsymbol{\alpha}^T \boldsymbol{\delta}(\mathbf{b}_i))^2 + \sum_{m=1}^{M_{\max}} \phi_m \sum_{\substack{\theta=1 \\ \theta^r = (\theta_1, \theta_2)}}^2 \sum_{\substack{r < s \\ r, s \in V(m)}} \int_{Q^m} \alpha_m^2 [D_{r,s}^\theta \delta_m(\mathbf{t}^m)]^2 dt^m, \quad (5)$$

Here, $\boldsymbol{\delta}(\mathbf{b}_i) := (1, \delta_1(\tilde{\mathbf{x}}_i^1), \delta_2(\tilde{\mathbf{x}}_i^2), \dots, \delta_m(\tilde{\mathbf{x}}_i^{M_{\max}}))^T$; $V(m) := \{ \kappa_j^m \mid j = 1, 2, \dots, K_m \}$ is the variable set related to the m th BF, δ_m ; $\mathbf{t}^m = (t_{m_1}, t_{m_2}, \dots, t_{m_{K_m}})^T$ presents the vector of variables contributing to the m th BF, δ_m ; $\phi_m \geq 0$ are *penalty parameters* ($m = 1, 2, \dots, M_{\max}$); $\boldsymbol{\alpha}$ is a $((M_{\max} + 1) \times 1)$ -parameter vector to be estimated by using the data points [16]. Furthermore, Q^m is some appropriate integration domains in K_m -dimensional parallel-pipe. After a discretization is employed to approximate the multi-dimensional integral

$\int \alpha_m^2 [D_{r,s}^\theta \delta_m(t^m)]^2 dt^m$ and only one penalty parameter ($\phi = \phi_m := \lambda^2$) is used rather than applying distinct penalty parameters for each derivative in (5), the new approximate form of PRSS can be written as follows (we refer to [2,7-9,16,20] for more details):

$$PRSS \approx \|y - \delta(\tilde{b})\alpha\|_2^2 + \phi \|L\alpha\|_2^2. \quad (6)$$

Here, L is to be assigned an $((M_{\max} + 1) \times (M_{\max} + 1))$ -diagonal matrix. Then, the PRSS problem looks like a classical RR problem with $\phi > 0, \phi = \lambda^2$ for some $\lambda \in \mathbb{R}$, and this problem in (6) is expressed as a CQP [21]. So, based on a suitable selection of the bound K , the optimization problem in (6) can be rearranged [16]:

$$\begin{aligned} & \underset{\alpha}{\text{minimize}} \quad t \\ & \text{subject to} \quad \|\zeta(\mathbf{b})\alpha - y\|_2 \leq t, \\ & \quad \quad \quad \|L\alpha\|_2 \leq \sqrt{K}. \end{aligned} \quad (7)$$

At this point, we state that a careful learning process has to be followed for the choice of K .

4. Natural gas demand modeling by RR and CMARS

In this study, we consider two multi-objective models which use ℓ_2 norm regularization in linear and nonlinear modeling for the prediction of natural gas demand. Here, the advantage of regularization is to decrease the risk of overfitting, which usually occurs in high-dimensional learning. The primary goal of the regularization technique is to make the machine learning algorithm “learn” but “not memorize” by penalizing the algorithm to decrease its generalization error to avoid the risk of overfitting. As a consequence, the variance of the model may be considerably declined, without losing any important properties in the data. Moreover, since regularization is a kind of robustification, these kinds of models can also be called robust models.

4.1. Data

In this study, the data set comprising of daily natural gas demand data from 2004 to 2013 are provided by Baskentgaz, the LDC of Ankara. To check the performance of our models, we draw on the validation technique. We divide the dataset into two subsets as training and testing sets. Here, the dataset is not divided randomly since it includes a time series of natural gas and meteorological variables. Instead, the first six years (from 2004 to 2009) of each variable under consideration are selected as the training dataset, while the last four years of the series are selected as the test dataset. In this study, demand for residential customers

of Baskentgaz is modeled by applying RR and CMARS algorithms.

Our dataset contains some meteorological variables namely heating degree day (average temperature), relative humidity, wind speed, daily maximum and minimum temperatures. In this study, the meteorological variables used are proved by the Turkish State Meteorological Service. Indeed, energy consumption, especially natural gas, is highly dependent on weather conditions. If the temperature drops below a certain value of the heating threshold, households use more energy owing to the excessive need for space heating [1,2].

Here, firstly, the model details are given and then their performances are discussed for each forecasting time horizon. In the application of RR, the MATLAB regularization toolbox is utilized. In applications of CMARS, Salford MARS [19] is utilized to obtain BFs for the large model provided by the forward MARS algorithm. Afterward, MOSEK optimization software and MATLAB are used to solve the CQP problem and estimate unknown parameters.

4.2. Criteria for performance evaluations

Here, we tested the prediction accuracy of two specific prediction methods on real-time data. The main performance indicator for checking the accuracy of the models is the mean absolute percentage error (MAPE). In addition to MAPE, we also evaluate the multiple coefficient of determination (R^2), correlation coefficient (r), average absolute error (AAE), and root mean square error (RMSE) to check the performances of proposed models. These measures and their formulas are presented in Table 1.

Table 1. Performance measures and their formulas.

Abbreviation	Formula
MAPE	$MAPE := \frac{1}{N} \sum_{k=1}^N \left \frac{y_k - \hat{y}_k}{y_k} \right $
R^2	$R^2 := 1 - \frac{\sum_{k=1}^N (y_k - \hat{y}_k)^2}{\sum_{k=1}^N (y_k - \bar{y}_k)^2}$
AAE	$AAE := \frac{1}{N} \sum_{k=1}^N y_k - \hat{y}_k $
RMSE	$RMSE := \sqrt{\frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2}$
r	$r := \frac{\sum_{k=1}^N (y - \bar{y})(\hat{y} - \bar{\hat{y}})}{\sqrt{s(y)^2 s(\hat{y})^2}}$

4.3. RR models

In the RR algorithm, first, many different models were obtained based on different penalty parameters, ϕ , by using the MATLAB regularization toolbox. Here, the penalty parameter, ϕ , controlled the relative effect of both criteria (bias and variance) on the estimation of regression coefficient in (1). Then, the penalty parameter that tries to minimize two criteria in a balanced manner was selected. So, the following RR model based on selected penalty value provided the best solution for complexity and accuracy in (1).

4.3.1. One-day-ahead prediction

Day-ahead forecast is generally used to reduce operational costs and eliminate the drawbacks that can occur owing to the imbalance between supply and demand quantities. The system operator requires to identify the supply security issues that may exist the next day. The RR model for one-day-ahead forecasting can be represented as follows:

$$\begin{aligned}\hat{Y}_7 = & 0.22 + 0.91X_1 - 0.11X_2 - 0.01X_3 + 0.01X_4 \\ & - 0.04X_5 + 0.04X_6 + 0.02X_7 + 0.09X_8 + 0.09X_9 \\ & - 0.03X_{10} + 0.04X_{11} + 0.05X_{12}.\end{aligned}$$

Here, for RR and CMARS models based on one-day ahead and one-week ahead, $X_1, X_2, X_3, X_4, X_5, X_6, X_7$, and X_8 represents the first-order lagged, second-order lagged, third-order lagged, fourth-order lagged, fifth-order lagged, sixth-order lagged, seventh-order lagged and fourteenth-order lagged natural gas consumption, respectively. Hence, the RR and CMARS models obtained are “lag” models since they involve lagged dependent variables. Furthermore, X_9, X_{10}, X_{11} , and X_{12} are the heating degree days, maximum temperature, minimum temperature, and wind speed, respectively.

4.3.2. One-week-ahead prediction

One-week-ahead forecast is usually used for generating unit production schedules for the next week. This gives companies insight into generation and consumption trends as well as assisting them to plan their weekly generation schedules. The RR model for a one-week forecast can be stated as follows:

$$\begin{aligned}\hat{Y}_7 = & 0.007 + 0.061X_1 + 0.06X_2 + 0.06X_3 + 0.06X_4 \\ & + 0.06X_5 + 0.06X_6 + 0.06X_7 + 0.6X_8 + 0.16X_9 \\ & + 0.03X_{10} - 0.02X_{11} + 0.01X_{12}.\end{aligned}$$

Here, RR models find all variables to be significant as expected.

4.4. CMARS Models

In the CMARS algorithm, first of all, using original data, the MARS algorithm is run via Salford MARS [22] and many MARS models were developed changing the number of BF and interaction. After selecting the optimal model for MARS among obtained models, the set of BFs in the forward part of MARS in the optimal MARS model was taken for the CMARS model. After the greatest models are constructed with the selected BFs and the L matrices in (7) are obtained, the PRSS in (6) is reformulated as a CQP problem. Here, based on different \sqrt{K} values in (7), several different CQP models in (7) were solved individually via MOSEK [11]. Here, MOSEK applies an interior point algorithm [12,13] to treat the CQP problems CMARS yields many solutions. Finally, the CQP model that has the minimum value of approximate PRSS was selected and the unknown parameters were estimated for the CMARS model. Here, this point tries to minimize the criteria, $\|\zeta(\mathbf{b})\boldsymbol{\alpha} - \mathbf{y}\|_2$ and $\|\mathbf{L}\boldsymbol{\alpha}\|_2$, in a balanced manner and the chosen value provides the best solution for complexity and accuracy based on PRSS. Here, we should note that the knot values of BFs in the CMARS model are selected differently but extremely close to the corresponding input data to avoid non-differentiability in the optimization problem.

4.4.1. One-day ahead prediction

In the CMARS algorithm, using the MARS software [22], the highest degree of interactions and M_{\max} are found by the forward MARS algorithm. For this data set, M_{\max} is 23 and the highest degree of interaction is 2. So, the largest model constructed by the forward MARS stage includes the following BFs.

$$\begin{aligned}\psi_1(\mathbf{x}) &= \max\{0, X_1 - 6.249\}, \\ \psi_2(\mathbf{x}) &= \max\{0, 6.249 - X_1\}, \\ \psi_3(\mathbf{x}) &= \max\{0, X_9 - 8.001\}, \\ \psi_4(\mathbf{x}) &= \max\{0, 8.001 - X_9\}, \\ \psi_5(\mathbf{x}) &= \max\{0, X_{11} - 7.801\}, \\ \psi_6(\mathbf{x}) &= \max\{0, 7.801 - X_{11}\}, \\ \psi_7(\mathbf{x}) &= \max\{0, X_6 - 9.925\}, \\ \psi_8(\mathbf{x}) &= \max\{0, 9.925 - X_6\}, \\ \psi_9(\mathbf{x}) &= \max\{0, X_9 - 10.301\} \cdot \max\{0, 9.925 - X_6\}, \\ \psi_{10}(\mathbf{x}) &= \max\{0, 10.301 - X_9\} \cdot \max\{0, 9.925 - X_6\}, \\ \psi_{11}(\mathbf{x}) &= \max\{0, X_1 - 2.502\} \cdot \max\{0, 9.925 - X_6\}, \\ \psi_{12}(\mathbf{x}) &= \max\{0, 2.502 - X_1\} \cdot \max\{0, 9.925 - X_6\}, \\ \psi_{13}(\mathbf{x}) &= \max\{0, X_{11} - 5.601\} \cdot \max\{0, 6.249 - X_1\}, \\ \psi_{14}(\mathbf{x}) &= \max\{0, 5.601 - X_{11}\} \cdot \max\{0, 6.249 - X_1\}, \\ \psi_{15}(\mathbf{x}) &= \max\{0, X_6 - 7.336\} \cdot \max\{0, X_9 - 8.001\},\end{aligned}$$

$$\begin{aligned}
\psi_{16}(\mathbf{x}) &= \max\{0, 7.336 - X_6\} \cdot \max\{0, X_9 - 8.001\}, \\
\psi_{17}(\mathbf{x}) &= \max\{0, X_1 - 4.685\} \cdot \max\{0, X_9 - 8.001\}, \\
\psi_{18}(\mathbf{x}) &= \max\{0, 4.685 - X_1\} \cdot \max\{0, X_9 - 8.001\}, \\
\psi_{19}(\mathbf{x}) &= \max\{0, X_{11} - 8.901\} \cdot \max\{0, X_9 - 8.001\}, \\
\psi_{20}(\mathbf{x}) &= \max\{0, 8.901 - X_{11}\} \cdot \max\{0, X_9 - 8.001\}, \\
\psi_{21}(\mathbf{x}) &= \max\{0, X_1 - 8.402\} \cdot \max\{0, X_9 - 8.001\}, \\
\psi_{22}(\mathbf{x}) &= \max\{0, 8.402 - X_1\} \cdot \max\{0, X_9 - 8.001\}, \\
\psi_{23}(\mathbf{x}) &= \max\{0, X_9 - 0.001\} \cdot \max\{0, X_6 - 9.925\}.
\end{aligned}$$

CMARS model for the one-day-ahead forecast is stated as follows:

$$\begin{aligned}
\hat{Y} &= \alpha_0 + \sum_{m=1}^{23} \alpha_m \psi_m(\mathbf{x}) \\
&= 5.30 + 0.77\psi_1(\mathbf{x}) - 0.61\psi_2(\mathbf{x}) - 0.27\psi_3(\mathbf{x}) - 0.09\psi_4(\mathbf{x}) \\
&\quad + 0.10\psi_5(\mathbf{x}) - 0.01\psi_6(\mathbf{x}) - 0.19\psi_7(\mathbf{x}) - 0.08\psi_8(\mathbf{x}) \\
&\quad + 0.01\psi_9(\mathbf{x}) + \alpha_{10}\psi_{10}(\mathbf{x}) - 0.002\psi_{11}(\mathbf{x}) - 0.01\psi_{12}(\mathbf{x}) \\
&\quad - 0.02\psi_{13}(\mathbf{x}) - 0.01\psi_{14}(\mathbf{x}) + 0.01\psi_{15}(\mathbf{x}) - 0.02\psi_{16}(\mathbf{x}) \\
&\quad + 0.05\psi_{17}(\mathbf{x}) - 0.09\psi_{18}(\mathbf{x}) + 0.01\psi_{19}(\mathbf{x}) + 0.02\psi_{20}(\mathbf{x}) \\
&\quad - 0.06\psi_{21}(\mathbf{x}) + 0.07\psi_{22}(\mathbf{x}) - 0.02\psi_{23}(\mathbf{x}).
\end{aligned}$$

For the CMARS model based on one-day ahead forecasting, the first-order lagged natural gas consumption (X_1), the sixth-order lagged natural gas consumption (X_6), the heating degree days (X_9), and minimum temperature (X_{11}) are significant.

4.4.2. One-week ahead prediction

For this natural gas data set based on one-week ahead forecasting, M_{\max} is 27 and the highest degree of interaction is 2. So, the largest model constructed by the forward MARS algorithm includes the following BF's.

$$\begin{aligned}
\psi_1(\mathbf{x}) &= \max\{0, X_9 - 7.001\}, \\
\psi_2(\mathbf{x}) &= \max\{0, 7.001 - X_9\}, \\
\psi_3(\mathbf{x}) &= \max\{0, X_1 - 10.221\}, \\
\psi_4(\mathbf{x}) &= \max\{0, 10.221 - X_1\}, \\
\psi_5(\mathbf{x}) &= \max\{0, X_{10} - 20.801\}, \\
\psi_6(\mathbf{x}) &= \max\{0, 20.801 - X_{10}\}, \\
\psi_7(\mathbf{x}) &= \max\{0, X_{10} - 19.201\} \cdot \max\{0, 10.221 - X_1\}, \\
\psi_8(\mathbf{x}) &= \max\{0, 19.201 - X_{10}\} \cdot \max\{0, 10.221 - X_1\}, \\
\psi_9(\mathbf{x}) &= \max\{0, X_1 - 2.217\} \cdot \max\{0, X_9 - 7.001\}, \\
\psi_{10}(\mathbf{x}) &= \max\{0, 2.217 - X_1\} \cdot \max\{0, X_9 - 7.001\}, \\
\psi_{11}(\mathbf{x}) &= \max\{0, X_3 - 1.572\} \cdot \max\{0, 10.221 - X_1\},
\end{aligned}$$

$$\begin{aligned}
\psi_{12}(\mathbf{x}) &= \max\{0, 1.572 - X_3\} \cdot \max\{0, 10.221 - X_1\}, \\
\psi_{13}(\mathbf{x}) &= \max\{0, X_8 - 6.032\} \cdot \max\{0, 10.221 - X_1\}, \\
\psi_{14}(\mathbf{x}) &= \max\{0, 6.032 - X_8\} \cdot \max\{0, 10.221 - X_1\}, \\
\psi_{15}(\mathbf{x}) &= \max\{0, X_{12} - 1.601\} \cdot \max\{0, X_9 - 7.001\}, \\
\psi_{16}(\mathbf{x}) &= \max\{0, 1.601 - X_{12}\} \cdot \max\{0, X_9 - 7.001\}, \\
\psi_{17}(\mathbf{x}) &= \max\{0, X_{12} - 1.601\} \cdot \max\{0, 20.801 - X_{10}\}, \\
\psi_{18}(\mathbf{x}) &= \max\{0, 1.601 - X_{12}\} \cdot \max\{0, 20.801 - X_{10}\}, \\
\psi_{19}(\mathbf{x}) &= \max\{0, X_3 - 10.937\}, \\
\psi_{20}(\mathbf{x}) &= \max\{0, 10.937 - X_3\}, \\
\psi_{21}(\mathbf{x}) &= \max\{0, X_1 - 1.733\} \cdot \max\{0, 10.937 - X_3\}, \\
\psi_{22}(\mathbf{x}) &= \max\{0, 1.733 - X_1\} \cdot \max\{0, 10.937 - X_3\}, \\
\psi_{23}(\mathbf{x}) &= \max\{0, X_6 - 6.287\} \cdot \max\{0, X_9 - 7.001\}, \\
\psi_{24}(\mathbf{x}) &= \max\{0, 6.287 - X_6\} \cdot \max\{0, X_9 - 7.001\}, \\
\psi_{25}(\mathbf{x}) &= \max\{0, X_7 - 3.521\} \cdot \max\{0, X_9 - 7.001\}, \\
\psi_{26}(\mathbf{x}) &= \max\{0, 3.521 - X_7\} \cdot \max\{0, X_9 - 7.001\}, \\
\psi_{27}(\mathbf{x}) &= \max\{0, X_{10} + 7.401\} \cdot \max\{0, 7.001 - X_9\}.
\end{aligned}$$

CMARS model for the one-week-ahead forecast is stated as follows:

$$\begin{aligned}
\hat{Y} &= \alpha_0 + \sum_{m=1}^{23} \alpha_m \psi_m(\mathbf{x}) \\
&= 5.24 + 0.43\psi_1(\mathbf{x}) + 0.65\psi_2(\mathbf{x}) + 0.44\psi_3(\mathbf{x}) + 1.07\psi_4(\mathbf{x}) \\
&\quad - 0.20\psi_5(\mathbf{x}) + 0.04\psi_6(\mathbf{x}) + 0.01\psi_7(\mathbf{x}) - 0.02\psi_8(\mathbf{x}) \\
&\quad - 0.01\psi_9(\mathbf{x}) + \alpha_{10}\psi_{10}(\mathbf{x}) - 0.18\psi_{11}(\mathbf{x}) - 0.18\psi_{12}(\mathbf{x}) \\
&\quad + 0.04\psi_{13}(\mathbf{x}) - 0.003\psi_{14}(\mathbf{x}) - 0.05\psi_{15}(\mathbf{x}) - 0.03\psi_{16}(\mathbf{x}) \\
&\quad + 0.04\psi_{17}(\mathbf{x}) + 0.06\psi_{18}(\mathbf{x}) - 0.15\psi_{19}(\mathbf{x}) - 1.36\psi_{20}(\mathbf{x}) \\
&\quad + 0.15\psi_{21}(\mathbf{x}) - 0.24\psi_{22}(\mathbf{x}) + 0.004\psi_{23}(\mathbf{x}) + 0.002\psi_{24}(\mathbf{x}) \\
&\quad - 0.04\psi_{25}(\mathbf{x}) + 0.02\psi_{26}(\mathbf{x}) - 0.02\psi_{27}(\mathbf{x}).
\end{aligned}$$

For the CMARS model based on one-week ahead forecasting, the first-order lagged natural gas consumption (X_1), the third-order lagged natural gas consumption (X_3), the sixth-order lagged natural gas consumption (X_6), the seventh-order lagged natural gas consumption (X_7), the fourteenth-order lagged natural gas consumption (X_{14}), the heating degree days (X_9), maximum temperature (X_{10}), and wind speed (X_{12}) are significant.

4.5. Results

In this study, we propose to model seasonal patterns and trends directly without using any transformations. As shown in Figures 1-8, the presented models capture the seasonality and trend of natural gas consumption. Moreover, with an additional plot, Figures 1-8 display the daily error (residual) time series for each model and let us analyze the residual dynamics. For natural gas demand prediction, the performance of the RR and CMARS models was evaluated using the performance measures presented in Table 1. For the training and test cases of the RR and CMARS models, Table 2 and Table 3 show the performance matrices for the forecasting of one-day-ahead and one-week-ahead, respectively.

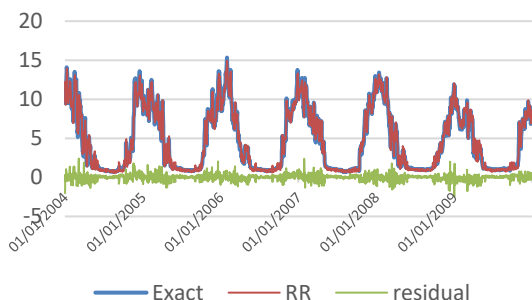


Figure 1. Exact values and RR approximation of natural gas consumption for one-day-ahead prediction based on training data

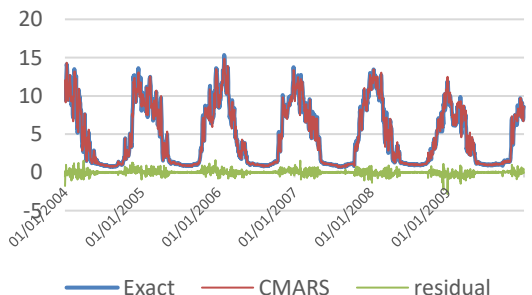


Figure 2. Exact values and CMARS approximation of natural gas consumption for one-day ahead prediction based on training data

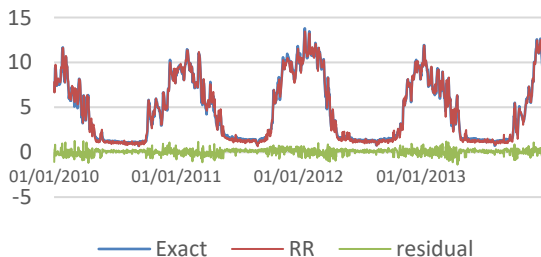


Figure 3. Exact values and RR approximation of natural gas consumption for one-day ahead prediction based on test data

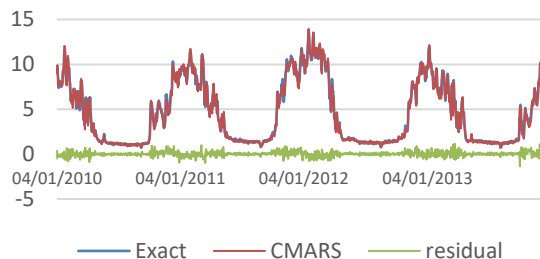


Figure 4. Exact values and CMARS approximation of natural gas consumption for one-day ahead prediction based on test data

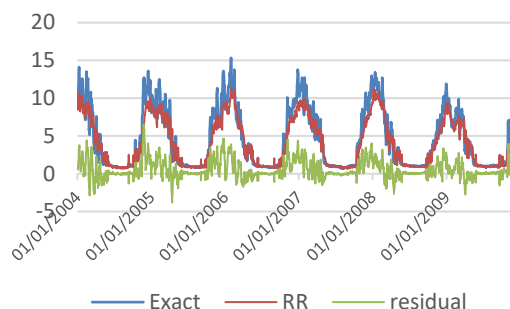


Figure 5. Exact values and RR approximation of natural gas consumption for one-week ahead prediction based on training data

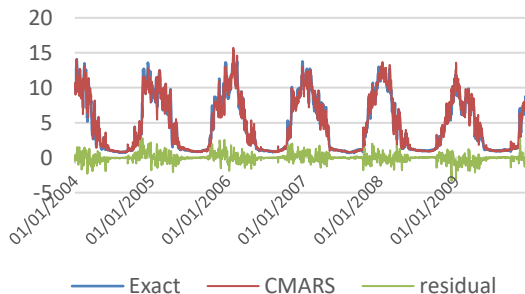


Figure 6. Exact values and CMARS approximation of natural gas consumption for one-week ahead prediction based on training data

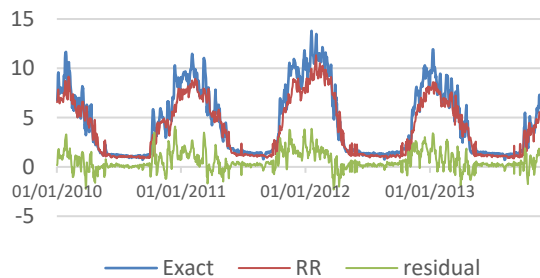


Figure 7. Exact values and RR approximation of natural gas consumption for one-week ahead prediction based on test data

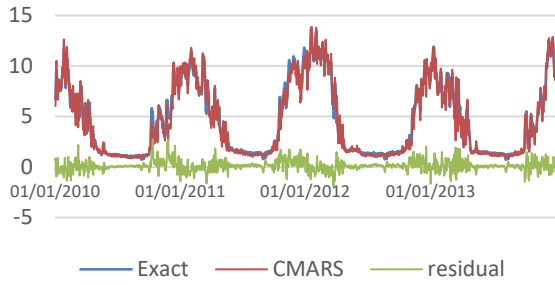


Figure 8. Exact values and CMARS approximation of natural gas consumption for one-week ahead prediction based on test data

Table 2. Performance of RR and CMARS for one-day ahead forecasting

	RR		CMARS	
	train	test	train	test
R^2	0.989	0.991	0.994	0.994
AAE	0.28	0.253	0.195	0.18
$RMSE$	0.405	0.333	0.313	0.264
r	0.995	0.996	0.997	0.997
$MAPE$	0.094	0.085	0.046	0.048

Table 3. Performance of RR and CMARS for one-week ahead forecasting

	RR		CMARS	
	train	test	train	test
R^2	0.885	0.889	0.977	0.976
AAE	0.861	0.827	0.383	0.368
$RMSE$	1.324	1.165	0.596	0.538
r	0.965	0.971	0.988	0.989
$MAPE$	0.181	0.183	0.09	0.099

For the one-day-ahead forecast, Table 2 shows the performances of the RR and CMARS models based on training and test. Depending on the results given in Table 2, the CMARS model outperforms the RR model in terms of all performance measures for training and testing data.

For the one-week-ahead forecast, the performances of RR and CMARS models are compared in Table 3 in terms of training and test cases. Based on the results shown in Table 3, the CMARS model outperforms the RR model in terms of all performance measures for training and test cases. Here, the MAPE performance becomes worse as the prediction horizon increases.

Based on residual, RR and CMARS have similar plots for one-day-ahead forecast whereas CMARS has better results than RR for the one-week-ahead forecast as in terms of training and test cases, as seen in Figures 1-8. Here, the proposed models have very small residuals for one-day-ahead forecasting. However, when the forecasting horizon increases, the residuals of the models become worse as expected.

In this application, the most *essential variables* in the analyzed residential demand models are based on *temperature* since a huge volume of natural gas is needed for space heating in Ankara. Therefore, in cold weather, natural gas usage reduces virtually linearly as the temperature rises. As you can see in Figures 1-7, this event also constructs a large difference between the winter and summer periods of private natural gas demand. Among the presented models, the CMARS model significantly outperforms the RR model in terms of all the training and test case performance criteria presented in Table 1 based on the short-term forecast. Therefore, CMARS should be the preferred model for this particular problem based on the short-term natural gas forecast.

5. Conclusion

In this study, we present two innovative models to short-term natural gas forecasting problems in the energy market. For residential users of LDCs, we evaluate daily and weekly forecasts of natural gas demand with daily intervals. We produce out-of-sample forecasts and compare them to observed data in terms of test datasets to assess the models' prediction accuracy.

The prediction accuracy is assessed for each forecast horizon using the performance criteria listed in Table 1. The proposed models, as illustrated in Figures 1-8, can capture the natural gas demand trend and seasonal pattern. However, it is found that CMARS performs better for one-day-ahead and one-week-ahead forecasts with MAPE values of 4.8% and 9.9%, respectively. Moreover, using CMARS as a substitute for MARS provides an integrated representation of all parameter identification tasks as a model-based optimization model instead of a model-free problem. As a result, in the multicriteria regression models offered, CMARS models should be favored for this particular situation.

However, since the knot selection does not require for the linear parts in CMARS, as further study, Conic Generalized Partial Linear Model (CGPLM) [23] may also be added to the CMARS algorithms so that these semi-parametric algorithms may provide to decrease the complexity of CMARS. Moreover, the same study may be repeated with the robust counterparts being R(C)PLM [24,25] and R(C)MARS [26-31], and then the comparative results may be provided. In future researches, proposed models can be compared to time series models and many other traditional and recent methods.

Acknowledgments


The authors would like to thank the anonymous reviewers for their valuable comments. This work was supported by TUBITAK under the 2219-International Postdoctoral Research Scholarship Programme.

References

- [1] Özmen, A., (2021). Sparse regression modeling for short- and long-term natural gas demand prediction. *Annals Operations Research*, DOI: <https://doi.org/10.1007/s10479-021-04089-x>
- [2] Özmen, A., Yilmaz Y., & Weber, G.-W. (2018). Natural Gas Consumption Forecasting Model for Residential User. *Energy Economics*, 70, 357–381.
- [3] Aster, R.C., Borchers, B., & Thurber C.H. (2013). *Parameter Estimation and Inverse Problems*. Academic Press, Boston.
- [4] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction of statistical learning*, Springer.
- [5] Wheeler, D. (2009). Simultaneous Coefficient Penalization and Model Selection in Geographically Weighted Regression: The Geographically Weighted Lasso. *Environment and Planning A*, 39(3), 722-742.
- [6] Friedman, J.H. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 19, 1–67.
- [7] Taylan, P., Weber, G.-W., & Beck, A. (2007). A new approach to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology, *Optimization*. 56, 5-6, 675–698.
- [8] Taylan, P., Weber, G. W., & Özkurt, F. Y. (2010). A new approach to multivariate adaptive regression splines by using Tikhonov regularization and continuous optimization. *Top*, 18(2), 377-395.
- [9] Weber, G.-W., Batmaz, İ., Köksal, G., Taylan, P., & Yerlikaya-Özkurt, F. (2012). CMARS: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization. *Inverse Problems in Science and Engineering*, 20(3), 371-400.
- [10] Ayyıldız, E., Purutçuoğlu, V., & Weber, G. W. (2018). Loop-based conic multivariate adaptive regression splines is a novel method for advanced construction of complex biological networks. *European Journal of Operational Research*, 270(3), 852-861.
- [11] Roos, C., Terlaky, T., & Vial, J. (1997). *Interior Point Approach to Linear Optimization: Theory and Algorithms*, John Wiley & Sons, New York.
- [12] Roos, C., Terlaky, T., & Vial, J. (2006). *Interior Point Methods for Linear Optimization*. Springer Science, Heidelberg/Boston.
- [13] MOSEK; software available at <http://www.mosek.com>.
- [14] Tibshirani, R.(1996). *Regression shrinkage and selection via the lasso*. *Journal of the Royal Statistical Society B*, 267-288.
- [15] Hastie, T., Tibshirani, R. & Friedman, J.H. (2001). *The Element of Statistical Learning*. Springer Series in Statistics; New York, USA.
- [16] Özmen, A., Kropat, E., & Weber, G.-W. (2014b). Spline regression models for complex multi-modal regulatory networks. *Optimization Methods and Software (OMS)* 29(3), 515-534.
- [17] Kriner, M. (2007). *Survival Analysis with Multivariate Adaptive Regression Splines*, Dissertation, LMU Munchen: Faculty of Mathematics, Computer Science and Statistics.
- [18] Hastie, T. & Tibshirani, R.J. (1996). Discriminant analysis by Gaussian mixtures, *Journal of the Royal Statistical Society (Ser. B)* 58, 155–176.
- [19] Leathwicka, J.R., Elith, J., & Hastie, T. (1993). Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions, *Ecological Modelling* 199 (2), 188-196.
- [20] Kuter, S., Weber, G.-W., Özmen, A., & Akyürek, Z. (2014). Modern applied mathematics for alternative modelling of the atmospheric effects on satellite images, *Modeling, Optimization, Dynamics and Bioeconomy I, Springer Proceedings in Mathematics & Statistics*, 73, 469-485.
- [21] Gürbüz, B., & Sezer, M. (2020). Modified operational matrix method for second-order nonlinear ordinary differential equations with quadratic and cubic terms. *An International Journal of Optimization and Control: Theories & Applications (IJOCTA)*, 10(2), 218-225.
- [22] Salford Systems-Data Mining and Predictive Analytics Software: MARS software available at <http://www.salfordsystems.com>.
- [23] Weber, G.-W., Çavuşoğlu, Z., & Özmen, A. (2012). Predicting default probabilities in emerging markets by conic generalized partial linear models and their optimization. *Optimization*, 61(4), 443-457.
- [24] Özmen A., Weber, G.-W., & Kropat, E. (2012). Robustification of conic generalized partial linear models under polyhedral uncertainty. *International IFNA-ANS scientific Journal "Problems of Nonlinear Analysis in Engineering Systems"*, 2(38), 104-113.
- [25] Özmen, A., Weber, G.-W., Çavuşoğlu, Z., & Deferli, Ö. (2013). The new robust conic GPLM method with an Application to Finance: prediction of credit default. *Journal of Global Optimization (JOGO)*, 56 (2), 233-249.
- [26] Özmen, A. (2016). Robust optimization of spline models and complex regulatory networks - theory, methods and application. Contribution to Management Science, Springer.
- [27] Özmen, A., Weber, G.-W., Batmaz, İ., & Kropat, E. (2011). RCMARS: Robustification of CMARS with Different Scenarios under Polyhedral Uncertainty Set. *Communications in Nonlinear Science and Numerical Simulation (CNSNS)* 16 (12), 4780-4787.

- [28] Özmen, A., Weber, G.-W., & Karimov, A. (2013). RCMARS: A new optimization supported tool applied on financial data. *Pacific Journal of Optimization (PJO)* 9(3), 535-552.
- [29] Özmen, A. & Weber, G.-W. (2014). RMARS: Robustification of Multivariate Adaptive Regression Spline under Polyhedral Uncertainty. *Journal of Computational and Applied Mathematics (CAM)*, 259, 914-924.
- [30] Özmen, A., Batmaz, I., & Weber, G.-W. (2014). Precipitation Modeling by Polyhedral RCMARS and Comparison with MARS and CMARS. *Environmental Modeling & Assessment*, 19(5), 425-435.
- [31] Özmen, A., Kropat, E., & Weber, G.-W. (2017). Robust Optimization in Spline Regression Models for Multi-model Regulatory Networks under Polyhedral Uncertainty. *Optimization*, 66 (12), 2135-2155.

Ayşe Özmen studied Industrial Engineering and also Mathematics and Computer Science as a double major at Çankaya University. She earned her M.S. and Ph.D. degrees in Scientific Computing at the Institute of Applied Mathematics from Middle East Technical University (METU). Her master thesis was honored by the Best Thesis Award of METU in 2012. Then her Ph.D thesis was honored by the Serhat Ozyar Award in 2016. Then, she has been in Canada for 1.5 years as a Post-doctoral Scholar at the University of Calgary, Department of Mathematics and Statistics. Her research field is on Optimization, Network Modeling, Data Science, Machine Learning, and Mathematical Modeling and Programming, Energy Modeling. Dr. Özmen has participated in many research projects from different areas and (co-)authored lots of special issues and articles.

 <http://orcid.org/0000-0003-3393-3619>

An International Journal of Optimization and Control: Theories & Applications (<http://ijocta.balikesir.edu.tr>)



This work is licensed under a Creative Commons Attribution 4.0 International License. The authors retain ownership of the copyright for their article, but they allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in IJOCTA, so long as the original authors and source are credited. To see the complete license contents, please visit <http://creativecommons.org/licenses/by/4.0/>.