Hybrid Models Combining Neural Networks and Nonparametric Regression Models Used for Time Series Prediction

DURSUN AYDIN
Department of Statistics
Muğla University
48000 Kötekli / Muğla
TURKEY
duaydin@mu.edu.tr

MAMMADAGHA MAMMADOV
Department of Statistics
Anadolu University
26470 Tepebaşı / Eskişehir
TURKEY
mmammadov@anadolu.edu.tr

Abstract: - In this paper, we proposed the hybrid models whose components are nonparametric regression and artificial neural networks. Smoothing spline, regression spline and additive regression models are considered as the nonparametric regression components. Furthermore, various multilayer perceptron algorithms and radial basis function network model are regarded as the artificial neural networks components. The performances of the models have been compared for the number of cars produced in Turkey. The results obtained by experimental evaluations show that hybrid models proposed in this study have performed much better in comparison to hybrid models examined by others (see for example, [1] and [2]).

Key-Words: - Time series; Neural networks; Multilayer perceptrons; Radial basis function; Nonparametric regression; Smoothing spline; Regression spline; Additive regression model; Hybrid models.

1 Introduction
In order to forecast time series apart from Autoregressive integrated moving average (ARIMA) and artificial neural networks (ANN), a hybrid approach that uses ARIMA and ANN models together is recommended by Tseng et al.[3], and [1]. Experimental results with real data sets in paper Zhang [1] indicate that a hybrid methodology that combines both ARIMA and ANN models can be an effective way to improve forecasting accuracy achieved by either of the models used separately. In addition, Aslanargun et al. [2] demonstrated that hybrid models combines models with two nonlinear components have had the best performance for time series forecasting. Zhang [1] explains the reasons of using hybrid models in detail.

In recently, nonparametric regression methods have become a very useful tool for non-linear data such as time series (E. Ferreira et al, [4]). However, these approaches perform poorly when seasonality is present. To overcome this problem, two alternatives methods have been proposed in literature. In both approaches the trend is specified as nonparametric, but the seasonal component specification is different. First, we discussed a semi-parametric model where the parametric part is a dummy-variable specification for the seasonality. Secondly, we considered the seasonal component to be a smooth function of time and, the model falls within the class of additive models. The nonparametric regression models are discussed in detail by Wahba [5], Hardle [6], Green ad Silverman [7], Hastie and Tibshirani [8], Hardle et al. [9].

2 Methodology
This paper will use the number of cars produced in Turkey for model estimation and evaluation. Firstly, ANN approach is considered and then, the nonparametric regression models and hybrid models, one of their components is nonparametric regression and the other is artificial neural networks model, are discussed in the next sections.

2.1 The ARIMA Model
ARIMA models are usually used to predict a univariate time series. In these models, any observed value of the series in any time period is defined as the linear component of the several past observations and random
errors. The general form of ARIMA model is given by [10]

\[ ARIMA(p, d, q)(P, D, Q), \]

where \( p \) is the number of parameters of the autoregressive (AR) model, \( d \) is degree of difference, \( q \) is the number of parameters in the moving average (MA) model, \( P \) is the number of parameters in AR seasonal model, \( D \) is the seasonal degree of difference, \( Q \) is number of parameters in MA seasonal model, and \( s \) is the period of seasonality (\( s = 4 \) for quarterly data and \( s = 12 \) for monthly data).

### 2.2 The ANN Approach to Time Series Modeling

Multilayer perceptrons (MLPs) are used in a variety of problems, especially in forecasting. Back-propagation (BP) is the widespread approximation approach for training of the multi-layer feed-forward neural networks based on Widrow-Hoff training rule [11]; [12]. The main idea here is to adjust the weights and the biases that minimize the sum of square error by propagation the error back at each step. To minimize the sum of square error, different BP algorithms are constructed by applying different numeric optimization algorithms among gradient and Newton methods class (see in detail [11]; [12] [13]; [14]; [15]): Gradient descent and improved gradient descent with momentum methods; Conjugate Gradient (CG) algorithms (Fletcher-Reeves; Polak-Ribiere and Scaled Conjugate Gradients (SCG)); Quasi-Newton (QN) algorithms (Broyden, Fletcher, Goldfarb, and Shanno (BFGS); One-step secant (OSS); Levenberg-Marquart (LM)).

In the Radial Basis Function Networks (RBF), one hidden layer with required number of units is enough in order to model a function. The activations of hidden (radial) units are defined depending on the distance of the input vector and the center vector. Typically, the radial layer has exponential activation functions and the output layer a linear activation function.

Having only one hidden layer and making faster education than the MLP, can be taken as advantages of the RBF. As the linear modeling methods are more useful in output layers of RBF, the difficulties that occur about the local minimums in the MLP are removed.

The MLP is usually used for time series modeling and forecasting [16]. For a hidden layer network architecture \( n°p:1 \) (\( n \), number of inputs, \( p \), number of hidden units and 1, number of outputs), inputs are the observed values of \( nth \) previous time points and outputs (targets) are \((n+1)th \) observed value. When the network square error function is examined, it can be seen that ANN is a non-linear functions of previous observations \( (y_{t-1}, y_{t-2}, ..., y_{t-n}) \) to \( y_t \) future observations [1]:

\[ y_t = f(y_{t-1}, y_{t-2}, ..., y_{t-n}, w) + e_t, \]

where \( (y_{t-1}, y_{t-2}, ..., y_{t-n}) \) denote input values, \( y_t \) denotes target (or output) value, \( w \) denote the weights of the network, \( e_t \) denote the vector of network error at time point \( t \). The predicted \( \hat{y}_t \) is calculated as follows:

\[ \hat{y}_t = f(y_{t-1}, y_{t-2}, ..., y_{t-n}, w) . \]

If \( N \) number of \( y_1, y_2, ..., y_N \) observations are used for a time series and 1-step forward forecast is made, the number of training samples are \( N - n \). \( (y_1, y_2, ..., y_n) \) is taken as first input training sample and \( y_{n+1} \) is accepted as the target. The second training pattern will contain \( (y_2, y_3, ..., y_{n+1}) \) as inputs and \( y_{n+2} \) as the second target output. Finally, \( (y_N, y_{N+1}, y_{N+2}, ..., y_{N+1}) \) and \( y_N \) will be the last inputs pattern and target correspondingly.

In training procedure, with the help of different BP algorithms, the parameters (weights and biases) of the network is obtained by getting closer to the minimum value of sum of square error \( SSE = \sum_{t=1}^{N} (y_t - \hat{y}_t)^2 \).

### 2.3 The Nonparametric Regression Approach for Time Series Modeling

For time series prediction, we consider the following basic model form

\[ y(t_i) = s(t_i) + f(t_i) + e(t_i), i = 1, ..., n \]

or

\[ y_i = s_i + z_i + e_i, i = 1, 2, ..., n. \]

where \( t_i \)’s are knot points spaced in time interval \([a, b]\), \( y_i \) = \( y(t_i) \) denote observations for response variable, \( s(t_i) \) denotes the seasonal component, \( f(t_i) \) represents the trend, and \( e(t_i) \) indicates the terms of error with zero mean and common variance \( \sigma_e^2 \).

The function \( f \) is estimated as a smooth function in \([a, b]\), but the estimation of the function \( s \) is different due to seasonality. Therefore, we considered two alternative models called as semi-parametric and additive regression model [4].

The semi-parametric regression model is given in following way:

\[ y_i = \sum_{k=1}^{r} \beta_k D_k + f(t_i) + e_i, i = 1, ..., n \]

where \( r \) is the number of annual observations (\( r = 12 \), \( D_k \)’s are dummy variable that denotes the seasonal effects, and \( \beta_k \)’s are parametric coefficients. Dummy
variables are denoted by \( D_i = D^*_i - D^*_{i} \), where \( D^*_i = 1 \) if \( i \) observation correspond to the \( k \)th month of year, and \( D^*_i = 0 \) otherwise, for cancels the seasonal effects when a year is completed [4]. Eq. (3) in vector-matrix form can be expressed as

\[
y = DB + f + e
\]

where \( \beta = (\beta_1, \ldots, \beta_i, \ldots) \), \( y = (y_1, \ldots, y_i) \), \( f = (f(t_1), \ldots, f(t_n)) \), and \( D \) is the \( n \times (r-1) \) matrix, so that \( D^T = (D^*)^{r-1} \).

Model (3) is a semi-parametric model due to consist of parametric linear component and nonparametric component. The main purpose is to estimate the parameter vector \( \beta \) and function \( f \) at sample points \( t_1, \ldots, t_n \). For this aim, two estimation methods, called as smoothing spline and regression spline, have been considered ([5]; [7]; [8]; [9]; [17]).

In smoothing spline method (SSM), mentioned here the vector parameter \( \beta \) and the values of function \( f \) at sample points \( t_1, \ldots, t_n \) are estimated by minimizing the penalized residual sum of squares

\[
PSS(\beta, \Omega) = \sum_{i=1}^{n} [y_i - f(t_i)]^2 + \lambda \int |f''(u)|^2 \, du \tag{5}
\]

where \( f \in C^2[0,1] \) and \( a^T \) is the \( i \)th row of the matrix \( D \). When the \( \beta = 0 \), resulting estimator has the form \( \hat{f} = (\hat{f}(t_1), \ldots, \hat{f}(t_n)) = S_\lambda y \), where \( S_\lambda \) a known positive-definite (symmetric) smoother matrix that depends on \( \lambda \) and the knots \( t_1, \ldots, t_n \) (see, [5]; [7]; [17]).

For a pre-specified value of \( \lambda \) the corresponding estimators for \( f \) and \( \beta \) based on Eq. (4) can be obtained as follows:

\[
\hat{\beta} = (D^T D)^{-1} D^T y \tag{6}
\]

\[
\hat{f} = S_\lambda (y - D\hat{\beta}) \tag{7}
\]

where \( S_\lambda \) a smoother matrix and \( D = (I - S_\lambda) D \). See in detail ([4]; [7]; [17]):

Evaluate some criterion function (such as cross validation, generalized cross validation) and iterate changing \( \lambda \) until it is minimized.

Smoothing spline become less practical when sample size \( n \) is large, because it uses \( n \) knots. Regression spline method (RSM) is a more general approach to spline fitting. A regression spline is a piecewise polynomial function whose highest order nonzero derivative takes jumps at fixed “knots”. Generally, regression splines are smoothed by deleting nonessential knots. When the knots are selected, regression spline can be fitted by ordinary least squares. For further discussion on selection of knots, see study of [18] and [19].

In RMS, the function \( f \) is approximated by linear combinations of the base functions \( 1, t, t^2, t^3 \) and \( (t - k_j)^3, j = 1, 2, \ldots, K \). In this case, \( f \) in equality (3) is given by

\[
f(t_i) = f(t_i; y_i) = \gamma_0 + \gamma_1 y_i + \gamma_2 y_i^2 + \sum_{j=1}^{K} b_j (t_i - k_j)^3, \quad i = 1, \ldots, n \tag{8}
\]

where \( b_1, \ldots, b_K \) are independently and identically distributed (i.i.d) with \( N(0, \sigma^2) \), \( (t)_s = t \) if \( t > 0 \) and 0 otherwise and \( k_1 < \ldots < k_K \) are fixed knots \( (\min(t_i) \leq k_1 < \ldots < k_K \leq \max(t_i)) \).

By considering the decomposition in (8) of the function \( f \), the parameters \( \beta = k_1, k = 1, 2, \ldots, r-1; \gamma_0, \gamma_1, \gamma_2, \gamma_3 \) and \( b_j, j = 1, 2, \ldots, K \) are obtained by minimizing the penalized residuals sum of squares [19].

The smoothing parameter (penalty parameter \( \lambda \)) and the number of knots \( K \) must be selected in implementing the regression spline. However, \( \lambda \) plays a more essential role. See Ruppert [18] for a detailed discussion of the knot selection. The solution can be obtained in S-Plus.

There are situations in which a dummy variable specification does not capture all fluctuations because of existing any seasonal effect. A more general case for seasonal component in (1) has been considered as follows

\[
y_i = g(t_i) = i, \quad i = 1, \ldots, n \tag{9}
\]

where \( g \in C^2[a,b] \) is a smooth function. By substitution of the equations (9) in (1), \( y_i \) is obtained as

\[
y_i = g(t_i) + f(t_i) + e(t_i), \quad i = 1, \ldots, n \tag{10}
\]

where \( e_i = e(t_i)'s \) are the terms of random error with zero mean and constant variance.

The model presented in (10) is called as additive nonparametric regression model (ARM). Additive models are discussed by Hastie and Tibshirani, [8]. Estimator of the model (10) is based on minimum of the penalized residual sum of squares

\[
PSS(f, g) = \sum_{i=1}^{n} [y_i - f(t_i) - g(t_i)]^2 + \lambda \int |f''(u)|^2 \, du + \lambda \int |g''(u)|^2 \, du \tag{11}
\]

where the first term in (11) denotes the residual sum of the squares (RSS) penalizing the lack of fit, the second term multiplicant by \( \lambda_f \) denotes the roughness penalty for the \( f \), and the third term multiplicant by \( \lambda_g \) denotes the roughness penalty for \( g \). In vector and matrix notation, Eq. (11) can be written as

\[
PSS(f, g) = (y - f - g)^T (y - f - g) + \lambda_f K_f f + \lambda_g K_g g \tag{12}
\]

where \( K_f \) is a penalty matrix for \( f \) and \( K_g \) is a penalty matrix for \( g \). The estimators of \( f \) and \( g \) are defined as:
\[ \hat{f} = (I + \lambda_1 K_j)^{-1} (y - g) = S_{\lambda_1} (y - g) \] (12)
\[ \hat{g} = (I + \lambda_2 K_j)^{-1} (y - f) = S_{\lambda_2} (y - f) \] (13)

2.4 The hybrid methodology
Suppose observed are measurements \( y_t, t = 1, 2, ..., N \).

The forecasts of the combined model in hybrid methodology are defined as follows:
\[ \hat{y}_t = \hat{y}^1_t + \hat{y}^2_t \]
where superscripts denote the row number of the hybrid model, \( \hat{y}^1_t \) and \( \hat{y}^2_t \) are estimations of the appropriate first and second combined models at the time point \( t \).

Firstly, to obtain the forecast values \( \hat{y}^1_t \) of the first model at time point \( t \), the model with \( 1 \)-index is applied to the observation data \( y_t, t = 1, 2, ..., N \). If the first model contains \( m_1 \) input units, the forecast values of the first model are calculated as follows:
\[ \hat{y}^1_t = f_1(y_{t-1}, y_{t-2}, ..., y_{t-m_1}) \]
where \( f_1 \) is the function obtained from the first model.

After this stage, the input data are calculated as \( e^1_t = y_t - \hat{y}^1_t \) for the second model. In this case, the number of \( e^1_t \) will be \( N - m_1 \) for this model. If the second model contains \( m_2 \) input units, the number of \( \hat{y}^2_t \) forecast (improved residuals) values will be \( N - m_1 - m_2 \). In this case, the forecast values that are appropriate for the second model are calculated as follows:
\[ \hat{y}^2_t = f_2(e_{t-1}^1, e_{t-2}^1, ..., e_{t-m_2}^1) \]
where \( f_2 \) is the function obtained from the second model.

In case of the nonparametric of the first model, \( m_1 = 0 \) and the number of \( e^1_t \) units will be \( N \). If the second model is nonparametric, the number of \( \hat{y}^2_t \) forecast values will be \( N - m_1 \). The hybrid model with good performance is obtained by the model evaluation criteria for the forecasting.

3 Experimental Evaluations
In this section, the number of cars produced in Turkey is examined. Appropriate nonparametric regression models and hybrid models were chosen by doing experiments to make forecasts, and these models are also compared. In this study, STATISTICA Neural Networks, S-Plus, and R-Programs are used.

3.1 Data Set
The data set, the number of produced car in Turkey between January 1989 and December 2008, is taken from the central bank of the republic of Turkey (www.tcmb.gov.tr)[20]. The data set is divided into two parts for the use in training and forecasting. In the first part, 216 monthly data are taken into account for the January 1989–December 2006 period. These data are used in training to construct the models. In the second part, with the help of the models constructed in the first part, the performances of those models are calculated using the 24 monthly data for the January 2007–December 2008 period.

4.2 Choice of Appropriate ANN Models
The choice of the best ANN models depend on a comparison of statistics such as the MSE (RMSE), MAE, and MAPE. As the initial weight and bias values of the network were random, 150 replications were made for the same network structure, and the models giving the best forecasts were determined.

As the cars data in question included seasonality, after trying many neural networks with different numbers of input units, as expected, the number of input units was determined as 12. During these experiments, various neural network algorithms with single layer, with one or two hidden layers, MLP, and RBF models were applied on the sample data set. As the initial 12 data were lost because of the seasonality, 204 from the 216 data were used to adjust the weights. In the training stage of the network, data were divided into two parts: 132 of the 204 data were used for training and 72 data were used for validation. This division was used to restrict memorization of the network and provided for better forecasts ([11]; [12]):

<table>
<thead>
<tr>
<th>Table 1. The weights and biases of the MLP(12:7:1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thresh</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>1.1</td>
</tr>
<tr>
<td>2.2</td>
</tr>
<tr>
<td>2.3</td>
</tr>
<tr>
<td>2.4</td>
</tr>
<tr>
<td>2.5</td>
</tr>
<tr>
<td>2.6</td>
</tr>
<tr>
<td>2.7</td>
</tr>
<tr>
<td>3.1</td>
</tr>
</tbody>
</table>
The estimators of the regression model to estimate the regression spline. Secondly, we considered an additive nonparametric model to estimate the nonparametric components [8]. We used two nonparametric regression models. Firstly, we considered a semi-parametric model to estimate the parametric vector $\beta$, and nonparametric function $f$. The estimators of the $\beta$ and $f$ are obtained by using (5). We need to select the smoothing parameter $\lambda$ presented in (5). In practice, a value of the smoothing parameter can be chosen by specifying degrees of freedom ($df = trace(S^\lambda)$) for the nonparametric components [8]. Therefore, we used the $df$ in order to select the smoothing parameter $\lambda$ in smoothing spline. On the other hand, both the smoothing parameter $\lambda$ and the number of knots $K$ must be selected in implementing the regression spline. Secondly, we considered an additive regression model to estimate $f$ and $g$ in eq. (10). The estimators of $f$ and $g$ are obtained by using eq. (11). We need to select the soothing parameters called as $\lambda_1$ and $\lambda_2$ in (11). We select the both of the smoothing parameters by specifying the $df$. Observed the number of produced car for January 1989–December 2006 period and the produced car estimation results obtained by appropriate nonparametric models are given in Figure 1.

<table>
<thead>
<tr>
<th>1.1</th>
<th>1.10</th>
<th>1.11</th>
<th>1.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>0.12</td>
<td>0.36</td>
<td>0.42</td>
</tr>
<tr>
<td>0.29</td>
<td>0.81</td>
<td>0.90</td>
<td>0.23</td>
</tr>
<tr>
<td>0.46</td>
<td>0.64</td>
<td>0.75</td>
<td>0.15</td>
</tr>
<tr>
<td>0.24</td>
<td>0.60</td>
<td>0.95</td>
<td>0.85</td>
</tr>
<tr>
<td>-</td>
<td>0.69</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>0.31</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>0.89</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.57</td>
<td>0.79</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>0.57</td>
<td>0.35</td>
<td>0.28</td>
<td>0.85</td>
</tr>
<tr>
<td>0.28</td>
<td>0.35</td>
<td>0.31</td>
<td>0.85</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

Note: The row and column header numeric terminology first lists the layer, then the unit number within the layer. For example, 2.1 stands for unit 1 in layer 2

Among the MLP networks, the MLP(12:7:1) model showed the best performance. The CG algorithm indicated the best performance at the 24th epoch. A hyperbolic tangent function is applied in the hidden unit and the linear activation function is applied in the output unit. The weights and biases of the MLP(12:7:1) model are given in Table 1

4.3 Constructing Appropriate Nonparametric Models

We used two nonparametric regression models. Firstly, we considered a semi-parametric regression model to estimate the parametric vector $\beta$, and nonparametric function $f$. The estimators of the $\beta$ and $f$ are obtained by using (5). We need to select the smoothing parameter $\lambda$ presented in (5). In practice, a value of the smoothing parameter can be chosen by specifying degrees of freedom ($df = trace(S^\lambda)$) for the nonparametric components [8]. Therefore, we used the $df$ in order to select the smoothing parameter $\lambda$ in smoothing spline. On the other hand, both the smoothing parameter $\lambda$ and the number of knots $K$ must be selected in implementing the regression spline. Secondly, we considered an additive regression model to estimate $f$ and $g$ in eq. (10). The estimators of $f$ and $g$ are obtained by using eq. (11). We need to select the soothing parameters called as $\lambda_1$ and $\lambda_2$ in (11). We select the both of the smoothing parameters by specifying the $df$. Observed the number of produced car for January 1989–December 2006 period and the produced car estimation results obtained by appropriate nonparametric models are given in Figure 1.

4.4 Constructing the Appropriate Hybrid Models

In this study, we discussed hybrid models where components are nonparametric regression and ANN. In determining hybrid models whose first component is nonparametric regression, firstly, the nonparametric regression model was applied to the 216 real data, and then, the 216 residuals were obtained. At the next step, as the second component, ANN was applied to the 216 residuals data. According to the forecasts resulting from experiment, the models with the best performance among hybrid models, whose first component is nonparametric regression, are ARM&RBF (12:4:1), ARM&MLP (12:8:1) RSM&RBF (4:9:1) RSM&MLP (6:8:1) SSM&RBF (12:3:1) SSM&MLP (12:6:1) The weights and biases of the second components of these hybrid models have been computed, but they are not presented here since they would occupy very much place.

When these models are constituted, the initial 12 data are lost because of the seasonality. 204 from the 216 data are used to adjust the weights. 132 of the 204 data are used for training and 72 data are used for validation. After the training stage is finished, appropriate hybrid models are obtained by the test data explained in next section.

4.5 Comparisons of the Models

We used the test data to compare the performances of the hybrid and the others models. The performances of models are evaluated the criterion such as the mean square error (MSE), the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE), see in detail Carey and Rob
Furthermore, the graph of observed and forecasted values of the same models is given in figure 2.

Figure 2. Observed and forecasted values for the January 2007–December 2008 period.

As shown figure 2, the most prominent feature of the data is that in due to summer holidays the production in August drops significantly, and this is well predicted by SSM&MLP(12:6:1).

The performance values of the nonparametric regression, ANN and hybrid models are calculated for the test data. These performance values are presented in following table 4. According to the values of MSE, RMSE, MAE and MAPE, the SSM&MLP(12:6:1) hybrid model has indicated the best empirical performance.

Table 4. Performance values for models selected

<table>
<thead>
<tr>
<th>Models</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(1,1,1)(1,1,0)2</td>
<td>2.16E+ 0</td>
<td>14707.6 3</td>
<td>12345.0 0</td>
<td>33.00</td>
</tr>
<tr>
<td>MLP (12:7:1)</td>
<td>1.69E+ 08</td>
<td>13010.8 7</td>
<td>10197.4 9</td>
<td>26.55</td>
</tr>
<tr>
<td>RBF(6:9:1)</td>
<td>2.22E+ 08</td>
<td>14907.3 7</td>
<td>13349.6 3</td>
<td>32.62</td>
</tr>
<tr>
<td>SSM</td>
<td>1.85E+ 08</td>
<td>13609.1 5</td>
<td>10025.8 0</td>
<td>28.60</td>
</tr>
<tr>
<td>RSM</td>
<td>2.86E+ 08</td>
<td>16921.1 0</td>
<td>10353.0 0</td>
<td>33.50</td>
</tr>
<tr>
<td>ARM</td>
<td>2.33E+ 08</td>
<td>15255.5 2</td>
<td>11385.8 0</td>
<td>31.00</td>
</tr>
<tr>
<td>SSM&amp;MLP (12:6:1)*</td>
<td>1.2E+0 8</td>
<td>11032.2 2</td>
<td>7807.3</td>
<td>23.70</td>
</tr>
<tr>
<td>SSM&amp;RBF (12:3:1)</td>
<td>2.81E+ 09</td>
<td>53008.0 7</td>
<td>51335.7 9</td>
<td>99.54</td>
</tr>
<tr>
<td>RSM&amp;MLP (6:8:1)</td>
<td>2.3E+0 8</td>
<td>15157.1 7</td>
<td>8826.4</td>
<td>27.80</td>
</tr>
<tr>
<td>RSM&amp;RBF (4:9:1)</td>
<td>2.78E+ 08</td>
<td>16668.6 1</td>
<td>10207.2 1</td>
<td>32.65</td>
</tr>
<tr>
<td>ARM&amp;MLP (12:8:1)</td>
<td>2.2E+0 8</td>
<td>14922.1 7</td>
<td>10968.2 2</td>
<td>30.10</td>
</tr>
<tr>
<td>ARM&amp;RBF (12:4:1)</td>
<td>2.23E+ 08</td>
<td>14936.2 4</td>
<td>11131.0 65</td>
<td>30.32</td>
</tr>
<tr>
<td>RBF</td>
<td>2.73E+ 09</td>
<td>52264.6 5</td>
<td>49797.4 8</td>
<td>94.31</td>
</tr>
<tr>
<td>RBF (12:18:1)&amp;SSM</td>
<td>6.71E+ 07</td>
<td>81941.4 7</td>
<td>79833.0 7</td>
<td>162.1</td>
</tr>
<tr>
<td>RBF (12:18:1)&amp;ARM</td>
<td>5.89E+ 09</td>
<td>76774.3 3</td>
<td>73385.1 0</td>
<td>140.7 7</td>
</tr>
<tr>
<td>MLP</td>
<td>5.11E+ 08</td>
<td>22611.9 9</td>
<td>20233.2 7</td>
<td>39.36</td>
</tr>
<tr>
<td>MLP</td>
<td>5.57E+ 05</td>
<td>23597.7 21</td>
<td>21253.9 40</td>
<td>40.48</td>
</tr>
</tbody>
</table>

(*) indicates the model having best performance

5 Conclusions

It is known that hybrid models indicate very good performance in time series forecasting problems. Zhang [1] reported that hybrid models where a component is linear and the other is nonlinear have demonstrated a good performance in time series forecasting. Then, a study has been made by Aslanargun et al., [2] and found that usage of hybrid models, whose components are nonlinear, are more effective.

In this paper, in the time series forecasting problems based on cars data, we observed that hybrid models where components are ANN and nonparametric regression, make more better forecasting than hybrid models discussed in Aslanargun et al.,[2]. We noticed that hybrid models, whose first component is SSM and second component is MLP, have indicated very good performance. At the same time, the hybrid models, whose second component is particularly MLP, have denoted good results too.

As a result, our opinion is that, using hybrid models, whose components are nonparametric regression and ANN, can be useful in time series forecasting problems included seasonality and trend. As the first component, we propose to use the nonparametric regression models, especially on these types of hybrid models. Furthermore, MLP model can be used as second component.

References: