

Tourism demand modelling and forecasting using data mining techniques in multivariate time series: a case study in Turkey

Selçuk CANKURT^{1,*}, Abdülhamit SUBAŞI²

¹Department of Information Technologies, Faculty of Engineering and Information Technologies,
International Burch University, Sarajevo, Bosnia and Herzegovina

²College of Engineering, Effat University, Jeddah, Saudi Arabia

Received: 15.11.2013

Accepted/Published Online: 12.03.2015

Final Version: 20.06.2016

Abstract: In this study multiple linear regression, multilayer perceptron (MLP) regression, and support vector regression (SVR) are used to make multivariate tourism forecasting for Turkey. This paper is a comparative study of data mining techniques based on multivariate regression modelling with monthly data points to forecast tourism demand; it focuses on Turkey. Both MLP and SVR methods are widely employed in the variety forecasting problems. Most of the previous research on tourism forecasting used univariate time series or a limited number of variables with mostly yearly or quarterly, and rarely monthly frequencies. However, the application of data mining techniques for multivariate forecasting in the context of tourism demand has not been widely explored. This paper differs from earlier research in two ways: 1) it proposes multivariate regression modelling with monthly data points to forecast tourism demand; and 2) it focuses on Turkey by using a dataset with the most recently accumulated (between January 1996 and Dec 2013) 67 time series with respect to Turkey and its top 26 major tourism clients. Comparison of forecasting performances in terms of relative absolute error (RAE) and root relative squared error (RRSE) measurements shows that the SVR model, with $RAE = 12.34\%$ and $RRSE = 14.02\%$, gives a better performance. The results obtained in this study provide information for researchers interested in applying data mining techniques to tourism demand forecasting and help policy makers, government bodies, investors, and managers for their regularization, planning, and investments by way of accurate tourism demand forecasting.

Key words: Tourism demand forecasting, multivariate forecasting, multiple linear regression, artificial neural network, multilayer perceptron regression, support vector regression

1. Introduction

Demand forecasting in the tourism sector is of great economic value both for the public and private stakeholders. Because we do not have an opportunity to stock most tourism products, accuracy in demand forecasting plays an important role for the tourism sector in enhancing decision-making, management effectiveness, competitiveness, and sustainable economic growth [1–3].

The World Travel and Tourism Council states that the largest industry in the world is travel and tourism. In addition to becoming the biggest industry, the tourism sector has been also the world's largest employer since 1992 [4]. Turkey is one of the fastest growing destinations in the world, behind China and Russia, according to the World Tourism Organization [5].

*Correspondence: selcuk.cankurt@ibu.edu.ba

Turkey shows a rapid growth rate in terms of tourist arrivals and revenues. In terms of tourist arrivals and tourism receipts, respectively Turkey is ranked 6th and 11th among the top 20 international tourism destinations. The number of foreign tourist arrivals from 2.1 million in 1985 reached 35.7 million in 2012; earnings from tourism increased from 840 million USD to 28 billion USD over the same period, according to the World Tourism Organization [6].

The current work is intended to provide accurate tourism projection for Turkey. Since tourism is the biggest industry in the world, and Turkey is one of the biggest players in the tourism market, research with respect to tourism and Turkey produces great economic value and makes a significant contribution to the tourism sector. Turkey has a promising potential, not only because of its long history, rich cultural heritage, beautiful nature, four-season climate, and hospitable people, but also because of its modern tourism facilities, competitiveness in prices, and quality tourism services.

The most commonly used models in the time series analysis for forecasting are traditional statistical methods. The drawback of these models is that they are linear models. The relationship between the variables is not linear for most problems in real life [7] and using linear models for such problems is not efficient. Conventional statistical methods (like multiple linear regression (MLR)) are suitable for data exhibiting specific patterns like trends, seasonality, and cyclicity.

Artificial neural network (ANN) techniques can efficiently deal with imperfect and irregular data. The tourism sector is heavily influenced by uncertainty and fluctuations such as promotions or extreme crises [8]. One major application area of ANNs is forecasting [9,10]. The ANN method was introduced to tourism forecasting in the late 1990s, and later it was increasingly used to forecast demands for tourism [3,8,11,12]. The ANN seems to be a better method for tourist arrival forecasting than the conventional statistical methods [13].

In recent years, support vector regression (SVR) methods have also been used for forecasting problems, and they have been successfully applied in tourism demand forecasting [3,14]. With the introduction of the ϵ -insensitive loss function [15], the support vector machine (SVM) method has been extended to solve nonlinear regression estimation problems [16] and its scope has been extended to forecasting problems. SVR models provide more accurate results than the neural network models and traditional linear regression models [16,17].

Our brief review of the literature (which focused on studies related to tourism demand approaches) indicated that univariate time series models with statistical and data mining techniques were frequently used in tourism forecasting, while multivariate time series models with data mining techniques have not been investigated much. The main purpose of the current study was to create a multivariate model to examine the forecasting performance of multiple linear regression (MLR), multilayer perceptron (MLP) regression, and SVR in the context of international tourism demand for Turkey.

2. Data and modelling approaches

In tourism demand forecasting, we should select: 1) metrics, which identify how to measure tourism demand; 2) determinants or factors, which explain the tourism demand; and 3) data mining methods, which explain the relationship between the demand and its determinants [11,12].

2.1. Measurements and determinants of the tourism demand

Tourism demand is predominantly measured in terms of the number of tourists coming from an origin country to a destination country, the number of the nights tourists spend in the destination country, or the expenditures

by tourists from an origin country in the destination country [12]. In this study, the number of tourists with monthly frequency is selected as a metric to measure the tourism demand for Turkey.

Most of the studies held for forecasting tourism demand for Turkey have paid little attention to modelling the demand function properly and determining the main motives of households' tourism decisions. Most previous studies have focused primarily on the demand side determinants of tourism such as income and price measurements, but the supply side factors such as accommodation capacity in the hosting country have been ignored [4].

One of the most comprehensive lists of determinants that influence the decisions and motivations of tourists was proposed in Uysal [18] and includes a wide range of economic, social-physiological, and exogenous factors. In the selection of input variables, there are two main considerations: 1) the difficulty of establishing relationships between the different characterized variables (for example, because images of destinations (a social-physiological factor), natural disasters (an exogenous factor), and the harmonic consumer price index (HCPI) (an economic factors) differ in nature, it is difficult to develop a data mining model that is capable of measure the tourism demand by employing the different characterized variables); and 2) the limitations of data availability. We chose our input variables in light of the study by Uysal [18] by considering their monthly availability.

In the present study, the focus was primarily on tourism demand time series in terms of the number of foreign arrivals to Turkey between 1996 and 2013 from the top 26 ranked tourism clients of Turkey; secondarily, the number of ministry-licensed hotel beds in Turkey, the number of tourism agencies in Turkey, the wholesale price index of Turkey, the gold selling price, the HCPI, and the exchange rates for tourism clients of Turkey were included into multivariate regression models as environmental and economic time series that might affect foreign tourism demand for Turkey. As additions to the variables proposed by Uysal [18], three index variables (used to identify the month, year, and season of the data instances) were included in the data set. A full list of the variables is given in Section 2.2.

2.2. Data set and multivariate approach for tourism demand modelling

When we develop forecasting models based on interest in single series (e.g., the number of tourists coming to Turkey), we examine the past data of this single series (so-called historical data) at hand, and try to reveal underlying relationships between the interest of our series and changes. Then we use these patterns to predict the likely future path.

In the multivariate forecasting framework, we should identify as many as possible relevant series that are likely to affect the phenomenon of our interest to develop more comprehensive and accurate forecasting models. We examine many different past data series at hand, and try to find underlying relationships between the phenomenon of our interest and these series, so that, based on the relationship discovered by the data mining techniques, we try to predict the likely future path of our phenomenon of interest.

The following is a causal multivariate model:

$$y_{t+h} = f(x_{t1}, x_{t2}, \dots, x_{tN}). \quad (1)$$

$x_{t1}, x_{t2}, \dots, x_{tN}$ represents observations provided from different data series at time t , N denotes the number of data series, and y_{t+h} is the h step ahead forecast (so-called forecasting horizon) made in time t .

We have constituted our data set with 67 time series (66 input variables + 1 target variable). The output attribute is the number of tourists coming to Turkey by shifting 12 months (y_{t+12}), and the input attributes ($x_{t1}, x_{t2}, \dots, x_{tN}$) are the list of wholesale prices index of Turkey, US Dollar selling price, 1 ons gold London

selling price in USD, hotel bed capacity of Turkey, number of tourism agencies in Turkey, HCPIs of leading clients of Turkey (namely Austria, Belgium, Denmark, France, Germany, Greece, Italy, Netherlands, Norway, Poland, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States, the Czech Republic, Bulgaria, and Romania), number of the tourists coming from the leading clients of Turkey (namely Germany, Russia, France, Iran, Bulgaria, Georgia, Greece, Ukraine, Azerbaijan, Austria, Belgium, Denmark, Holland, England, Spain, Sweden, Switzerland, Italy, Norway, Poland, Romania, USA, Iraq, the Czech Republic, Kazakhstan, and Japan), exchange rate of the currencies of leading clients of Turkey (Danish Crone, Norwegian Crone, Polish Zloty, Swedish Crone, Swiss Franc, Turkish Lira, British Pound, Russian Ruble, Bulgarian Lev, Romanian Leu, Czech Koruna, and Japanese Yen), number of the former tourists and index of year, month, and season. All 66 time series used for input in the development of the multivariate tourism demand significantly influence the model performance. All values in the given dataset are normalized in the range of [0, 1].

Our input data, which form a matrix $\mathbf{X} \in R^{T \times N}$, and the respected outputs, which form a column vector $\mathbf{Y} \in R^{T \times 1}$, can be illustrated in a compact matrix format as:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{T1} & \cdots & x_{TN} \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} Y_{1+h} \\ \vdots \\ Y_{T+h} \end{bmatrix}$$

where $T=204$ is the number of data collection points in a time period, $N = 66$ is the number of the input variables (time series), and $h = 12$ is the forecasting horizon. x_{tn} is an instance of input data collected at time t , for the n^{th} variable. Y_{t+h} is the output response at time t . To map the input matrix \mathbf{X} to the output response vector \mathbf{Y} , in order to determine the functional relationship f (Eq. (1)), we used the MLR, MLP regression, and SVR algorithms.

Monthly time series data were collected from the Ministry of Tourism of the Republic of Turkey (website: www.turizm.gov.tr), The State Institute of Statistics of Turkey (website: www.die.gov.tr), The Databank of the Central Bank of the Republic of Turkey (website: <http://evds.tcmb.gov.tr>), The European Central Bank (ECB) (website: <https://www.ecb.europa.eu>), and TÜRSAB (website: www.tursab.org.tr).

2.3. MLR

MLR attempts to explain the linear relationship called the regression function between one dependent variable and more than one explanatory variables.

The model for MLR, given n observations, is:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_p x_{i,p} + \varepsilon_i \quad (2)$$

where $x_{i,p}$ is the value of p^{th} predictor, β_0 is the intercept or the bias in machine learning, β_p is the coefficient on the p^{th} predictor, p is the total number of predictors, y_i is the predictand, and ε_i is the error.

2.4. ANN approach

ANN is made up of highly interconnected computational processing units (so-called artificial neurons) inspired by biological nervous systems. They have the ability of learning by adjusting the strength of interconnections that can be achieved by altering the values known as synaptic weights through the input data [19]. The

processing unit (neuron) sums the weighted inputs and transfers the net input through an activation function in order to normalize and produce a result [20].

The multilayer feed-forward neural network is the most commonly used neural network architecture. Backpropagation (BP), which is based on the Widrow–Hoff training rule, is the most popular training algorithm for multilayer feed-forward neural networks [19,21]. There are two types of error correction functions for BP. The first error correction function (Eq. (3)) is employed for output neurons, and the second one is used only for hidden neurons (Eq. (4)) [13].

$$E_0 = (a_i - y_i) g'(y_i) \quad (3)$$

$$E_h = \left(\sum_{i=1}^{i < n} (w_{h,i} E_0) \right) g(y_h) \quad (4)$$

where y is the adjusted output of the i^{th} node, a is the expected result, w represents all of the weights (from 1 to n) connecting the hidden nodes to all inputs nodes, $g()$ is the transfer function, and g' is the first derivative of the activation function.

The next step is to adjust the corresponding weights for the node by using this error. We use Eq. (5) to update the weights, which uses the error previously calculated for the neuron (whether hidden or output):

$$w_{ij} = w_{ij} + \mu E y_i \quad (5)$$

where μ is the learning rate, which is a user-specified parameter. The weight is updated by multiplying the learning rate with the calculated error (E) and neuron output (y_i), and then adding this to the current weight. The result is a minimization of the error at this node, while the output node activation approaches the expected output [20].

2.5. Support vector regression

In supervised learning, SVM is a new type and currently the most popular learning algorithm. SVM tries to find a linear separating hyperplane by mapping data into the feature space with the higher-dimensional space using the so-called kernel trick. Even samples that are not linearly separable in the original input space can be easily separated in the higher-dimensional space by using the linear hyperplane [22,23]. SVM has been successfully employed for dealing with nonlinear regression problems and time series prediction applications [15].

Suppose we are given training data $\{(x_1, y_1), \dots, (x_i, y_i)\} \subset \mathfrak{N} \times \mathfrak{R}$, where \mathfrak{N} denotes the space of the input patterns. In ε -SV regression, the goal is to find a function $f(x)$ that has at most ε deviations from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible [22]. The case of the linear function $f(x)$ has been described in the following form:

$$f(x) = \langle w, x \rangle + b, \text{ with } w \in \mathfrak{N}, b \in \mathfrak{R} \quad (6)$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product in \mathfrak{N} . $\langle w, x \rangle$ is called feature, which is nonlinearly mapped from the input space x . The w and b are the coefficients, which are estimated by minimizing the regularized risk function. Flatness in the case of Eq. (6) means that one seeks a small w .

The learning procedure of SVM can be explained as follows. The minimization of complexity term can be achieved by minimizing the quantity defined in the following:

$$\frac{\|w\|^2}{2} \tag{7}$$

The coefficients of w and b can be estimated by minimizing the regularized risk function, as follows:

$$R(C) = \left(\frac{C}{N}\right) \sum_{i=1}^N L_\varepsilon(d_i, y_i) + \frac{\|w\|^2}{2}, \tag{8}$$

where

$$L_\varepsilon(d_i, y_i) = \begin{cases} 0 & |d - y| \leq \varepsilon \\ |d - y| - \varepsilon & \text{otherwise,} \end{cases} \tag{9}$$

$L_\varepsilon(d_i, y_i)$ is called the ε -insensitive loss function (Eq. (9)). C and ε are user-defined parameters in the empirical analysis. The parameter ε is the difference between the actual values and the values calculated from the regression function. Differences less than or in the size ε outline a space, so-called ε -tube, around the regression function [22,24,25].

2.6. Prediction performance metrics

In the regression, rather than determining whether the predicted value is right or wrong, we should consider how close or how far the predicted values are to the observed values [23]. We used root relative squared error (RRSE), relative absolute error (RAE), and correlation coefficient (R , sometimes also denoted r) measures to demonstrate the performance of the proposed models. These measures are defined below in Eq. (10), where n is the number of the test cases, a_i is the actual (observed) value for the i^{th} test case, and p_i is the predicted (estimated) value for the i^{th} test case:

$$\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i, \quad \bar{p} = \frac{1}{n} \sum_{i=1}^n p_i,$$

$$S_A = \frac{1}{n-1} \sum_{i=1}^n (a_i - \bar{a})^2, \quad S_P = \frac{1}{n-1} \sum_{i=1}^n (p_i - \bar{p})^2, \quad S_{PA} = \frac{1}{n-1} \sum_{i=1}^n (p_i - \bar{p})(a_i - \bar{a}) \tag{10}$$

Root relative squared error (RRSE)	Relative absolute error (RAE)	Correlation coefficient (R)
$RRSE = \sqrt{\frac{\sum_{i=1}^n (p_i - a_i)^2}{\sum_{i=1}^n (a_i - \bar{a})^2}}$	$RAE = \frac{\sum_{i=1}^n p_i - a_i }{\sum_{i=1}^n a_i - \bar{a} }$	$R = \frac{S_{PA}}{S_P S_A}$

The coefficient of determination (R^2) shows the percentage of data that can be explained by the regression equation. As R varies in the range of $-1 \leq R \leq 1$, R^2 gets the values in the range of $0 \leq R^2 \leq 1$ [23].

3. Results and discussion

Forecasting methods provide decision makers superior information for optimal decision-making over time. Analytic assessment of past data and forecasting are different activities, but are strongly related [26].

The purpose of this study was to investigate the applicability of data mining techniques such as MLR, MLP regression, and SVR in forecasting tourism demand for Turkey with respect to its 26 major clients by using multivariate time series among a wide range of the data series for a period between 1996 and 2013 with monthly collected historical data.

In order to achieve this goal, we developed many different data mining models on the basis of 3 regression models (MLR, MLP regression, and SVR) and then compared the best models in each category. Those models were evaluated and compared by using the same validation data through 3 forecasting accuracy measures: correlation coefficient (R), RAE, and RRSE.

In our implementations, we used WEKA [27] data mining software, which is an open source software issued under the GNU General Public License.

3.1. Evaluation of the models

We developed, implemented, and examined several different configurations of the MLR, MLP regression, and SVR models according to their corresponding parameter selection. Because of space consideration, only 29 of them (1 MLR, 14 MLP regression, and 14 SVR models) were reported in this paper.

We employed the 10-fold cross-validation approach to select the training and testing data sets. This approach is worthwhile to represent the characteristics of the full data set in about the right proportion in the training and testing sets when the amount of data for training and testing is limited [23]. In this process, the data set was randomly divided into 10 folds (subsets); each of the 10 folds contained approximately the same number of data points. In each turn, different 9/10 folds of the data set were used for building a model and the remaining 1 fold alone was used for testing the model. At the end, each instance in the dataset became one time a member of any one of the randomly selected 10 different test folds. Since we randomly generated 10 different test folds in the 10-fold cross-validation approach, each run generated 10 error estimates. The error estimates of RAE and RRSE and the R values for the 10 different test folds were averaged and reported.

3.2. Results

The first model that was employed through this study was based on the MLR method to make predictions. MLR models are implemented and examined on the basis of the attribute selection method; when no attribute selection is possible, the M5's method (eliminate the features with the smallest standardized coefficient until no more improvement is obtained in the estimate of the error given by the Akaike information criterion) and a greedy selection using the Akaike information metric are used [23]. Our investigation showed that MLR with attribute selection of the M5's method has good accuracy with $R = 0.9863$, $RAE = 14.86\%$, and $RRSE = 16.47\%$, but not better than the MLP regression and SVR models. The MLR with M5's attribute selection method generates a regression model to explain the statistical relationship between the predictor variables and the response variable in the given dataset. The features with the smallest standardized coefficient are eliminated [23]. The regression equation involves only predictor variables that significantly contribute to the accuracy of the regression model.

The generated regression model is: Total visitors(tourism demand) = $(124718.5763 \times \text{season index}) + (-0.0665 \times \text{whole sale prices of Turkey}) + (-0.3981 \times \text{US Dollar selling price}) + (541.3473 \times 1 \text{ Ons gold})$

London selling price) + (-26492.8282 × HCPI Belgium) + (-11457.9889 × HCPI Bulgaria) + (72606.8905 × HCPI Denmark) + (90467.7301 × CPI Germany) + (-11175.5536 × HCPI Greece) + (25805.8048 × HCPI Spain) + (-30788.0905 × HCPI France) + (48311.2161 × HCPI Italy) + (-101454.7032 × HCPI Austria) + (-27923.6335 × HCPI Poland) + (-2513.697 × HCPI Romania) + (-34746.889 × HCPI Sweden) + (28291.911 × HCPI United Kingdom) + (50598.6895 × HCPI Switzerland) + (33348.8821 × HCPI Turkey) + (-29844.1048 × HCPI United States) + (196793.5742 × Bulgarian Lev) + (-827562.5627 × Pound Sterling) + (133703.2873 × Romanian Leu) + (776552.4609 × Turkish Lira) + (-33386.3992 × Russian Ruble) + (62514.0433 × Czech Koruna) + (-2929.6509 × Japanese Yen) + (0.6795 × German visitors) + (0.793 × Russian visitors) + (1.1632 × British visitors) + (1.513 × Iranian visitors) + (-1.4107 × Bulgarian visitors) + (2.1566 × Dutch visitors) + (4.0437 × Italian visitors) + (7.2466 × Ukraine visitors) + (1.9152 × Belgian visitors) + (5.5436 × Polish visitors) + (6.8262 × Iraqi visitors) + (-4.8501 × Norwegian visitors) + (-5848254.1722).

To train the multilayer feed-forward neural networks, we employed BP algorithm, which is based on the Widrow–Hoff training rule [19,21]. The sigmoid transfer function is preferred due to its significant achievement to produce models with sufficient accuracy [28]. We used the sigmoid transfer function for the hidden nodes, and the unthresholded linear activation function for the output node.

The hidden layer of an ANN model acts as a black box to link the relationship between input and output [28]. One of the applied results of Kolmogorov’s theorem for neural networks states that 2 hidden layers are enough for a certain approximation of any complex nonlinear function. In fact, usually 1 layer in a network is satisfactory in order to construct an approximation function [10,13,21]. There is no rule that indicates the optimum number of hidden neurons for any given problem. According to prior studies, the number of neurons in the hidden layer can be up to (1) $2n + 1$ (where n is the number of neurons in the input layer), (2) 75% of input neurons, or (3) 50% of input and output neurons [8,29,30].

The learning rate controls the amount of changes in weight and reduces the possibility of any weight oscillation during the training cycle. A learning rate between 0.05 and 0.5 provides good results in most practical cases [31]. The momentum factor determines the effect of past changes and current changes in weights and speeds up the learning time. The momentum factor usually has a value of close to 1, e.g., 0.9 [31,32]. The network stops when a specified number of epochs or the learning rate is reached. Training continues until performance on the validation set is satisfactory or until the specified number of epochs is reached [23].

Among the presented MLP regression models, the MLP model ID 6 had the best forecasting accuracy (Table 1); it was composed of 3 hidden layers with neuron numbers of 33, 10, and 6 (abbreviated in the table as 33–10–6), with learning rate of 0.01, momentum of 0.7, and an epoch value of 500. It can be seen from Table 1 that MLP model ID 6 with $R = 0.9875$, $RAE = 14.19\%$, and $RRSE = 15.82\%$ achieved the best performance among the other MLP regression models. Moreover, in Figures 1 and 2 the comparison and correlation, respectively, between the actual values and the corresponding values predicted by MLP model ID 6 can be seen.

The SVR models used in this study are based on the sequential minimal optimization algorithm, which was originally proposed by Platt [33], improved by Schölkopf and Smola [34], and extended by polynomial or Gaussian kernels [33] for SVR problems [23,35].

Because of the SVR model’s complexity, selection of the optimal parameter is further complicated. To construct the SVR model efficiently, SVR parameters must be set carefully [36–38]. Inappropriate parameters

in SVR lead to overfitting or underfitting [14,38]. The most important parameters for the SVR model are the type of kernel function that is used to model a nonlinear decision hypersurface on the SVR input space and the regularization parameter (C), which controls the trade-off between the training error and the complexity. However, there is no general systematic method to select parameters used in the SVR model [22,34, 39–41].

Table 1. Forecasting performance of the MLP-based models for Turkey.

MLP model ID	Learning Rate	Momentum	# of neurons in the hidden layers*	The number of epochs	Correlation coefficient	Relative absolute error	Root relative Squared Error
1	0.01	0.2	4	500	0.9867	14.20%	16.30%
2	0.03	0.8	33–11	500	0.9858	14.80%	16.91%
3	0.01	0.5	33–11	500	0.9866	14.41%	16.37%
4	0.01	0.8	66–33–11	500	0.9866	14.46%	16.38%
5	0.01	0.8	67–33–11	500	0.9866	14.48%	16.36%
6	0.01	0.7	33–10–6	500	0.9875	14.19%	15.82%
7	0.01	0.7	67–11	500	0.9853	14.70%	17.21%
8	0.01	0.8	67–67	500	0.9825	15.34%	18.87%
9	0.01	0.7	67–67	500	0.9827	15.60%	18.92%
10	0.01	0.7	66–67	500	0.9823	15.65%	19.20%
11	0.01	0.8	66–33	500	0.9832	15.13%	18.45%
12	0.01	0.8	67–20	500	0.9845	14.88%	17.70%
13	0.01	0.8	33–15	500	0.9851	14.68%	17.31%
14	0.01	0.8	67–33–67	500	0.9864	14.75%	16.43%

*67 = # of attributes + class, 66 = # of attributes.

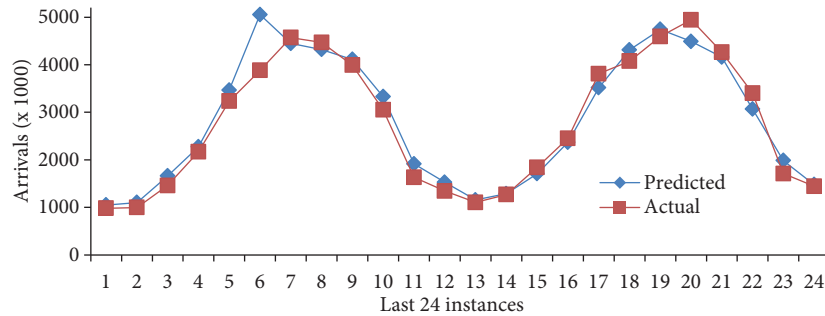


Figure 1. Comparison of actual values with the results obtained from MLP model ID 6.

The results for the 14 SVR models are reported in Table 2. SVR models are implemented and examined on the basis of variation in the following parameters: complexity (C), kernel function, and the kernel function exponent, gamma, omega, and sigma parameters. It can be seen from Table 2 that among the SVR models using the polynomial kernel, model ID 2, with complexity parameter 0.1 and exponent parameter 2, showed better accuracy with $R = 0.9862$, $RAE = 13.59\%$, and $RRSE = 16.47\%$. Among the SVR models using the normalized polynomial kernel, model ID 6, with complexity parameter 2 and exponent parameter 3, showed better accuracy with $R = 0.9899$, $RAE = 12.82\%$, and $RRSE = 14.13\%$. Among the SVR models using the RBF kernel, model ID 10, with complexity parameter 25 and gamma parameter $\gamma = 0.02$, showed better accuracy with $R = 0.9897$, $RAE = 12.49\%$, and $RRSE = 14.29\%$. Among the SVR models using the PUK kernel, model ID 13 with complexity parameter $c = 16$, omega parameter $\omega = 3$, and sigma parameter $\sigma = 11$,

showed the best accuracy with $R = 0.9901$, $RAE = 12.34\%$, and $RRSE = 14.02\%$. This model (among the models evaluated in this study) gives also the best results overall. Table 2 also shows that the forecasting performance of the SVR model with the PUK kernel among the 14 SVR models was superior to the ones with the other kernels. Comparison of the actual tourist arrivals with the forecasted tourist arrivals by SVR model ID 13 is shown in Figure 3. In addition, Figures 4 and 5 show the performance of this model by illustrating correlation and error bars between the actual values and the corresponding predicted values.

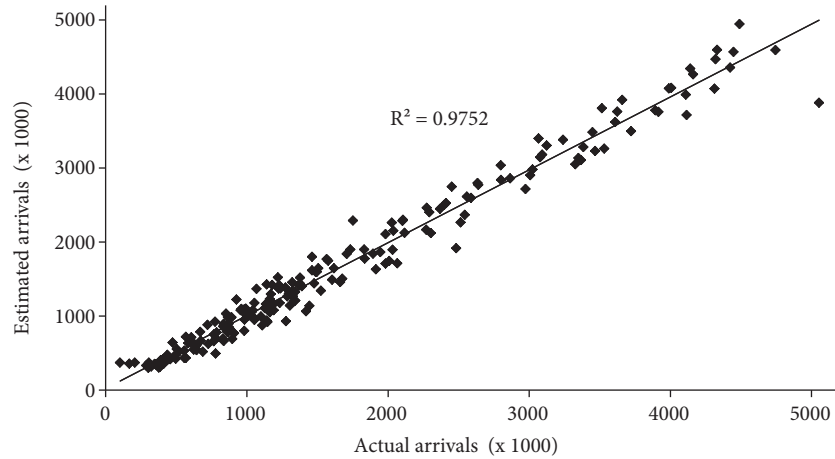


Figure 2. Performance of MLP model ID 6 for a dataset.

Table 2. Forecasting performance of the SVR-based models for Turkey.

SVR model ID	C	Kernel type	Correlation coefficient	Relative absolute error	Root relative squared error
1	2	The polynomial kernel (Exponent: 1)	0.9758	16.15%	22.50%
2	0.1	The polynomial kernel (Exponent: 2)	0.9862	13.59%	16.47%
3	0	The polynomial kernel (Exponent: 2)	0.9874	13.64%	15.80%
4	0	The polynomial kernel (Exponent: 3)	0.9718	16.56%	23.51%
5	2	normalized polynomial kernel (Exponent: 4)	0.9895	13.20%	14.48%
6	2	normalized polynomial kernel (Exponent: 3)	0.9899	12.82%	14.13%
7	1	normalized polynomial kernel (Exponent: 2)	0.9885	13.17%	15.15%
8	10	The RBF kernel (Gamma: 0.04)	0.9878	12.60%	15.57%
9	10	The RBF kernel (Gamma: 0.03)	0.989	12.65%	14.76%
10	25	The RBF kernel (Gamma: 0.02)	0.9897	12.49%	14.29%
11	100	The RBF kernel (Gamma: 0.01)	0.9893	12.56%	14.53%
12	10	PUK (Omega/Sigma: 3/11)	0.99	12.47%	14.10%
13	16	PUK (Omega/Sigma: 3/11)	0.9901	12.34%	14.02%
14	20	PUK (Omega/Sigma: 3/6)	0.9898	12.49%	14.25%

To evaluate how accurate the results of the developed models are, the coefficient of correlation (R^2) was also used as a statistical verification tool. Estimated values were graphically correlated with the actual values as in Figures 2 and 4. The models were found to be able to learn the relationships among the input parameters, i.e. the number of tourists, the exchange rates, HCPIs, wholesale prices index, gold price, hotel bed capacity, number of tourism agencies. Figure 2 gives the statistical performance of the ANN model. It appears that there is a relatively good agreement between the ANN predictions and the actual data. This can be inferred from the R^2 value of 0.9752 (Figure 2). Figure 4 gives the statistical performance of the SVR model. Similarly, there is

a relatively good agreement between the SVR predictions and the actual data. This can be inferred from the R^2 value of 0.9803 (Figure 4). The R^2 value of the model reflects the overall error performance of the model. One can clearly see that the SVR model gives better correlation between the estimated and the real tourism demand for Turkey than the ANN model. Consequently, when the results in the Figures are evaluated, it can be concluded that we can use the SVR model for the prediction of tourism demand for Turkey.

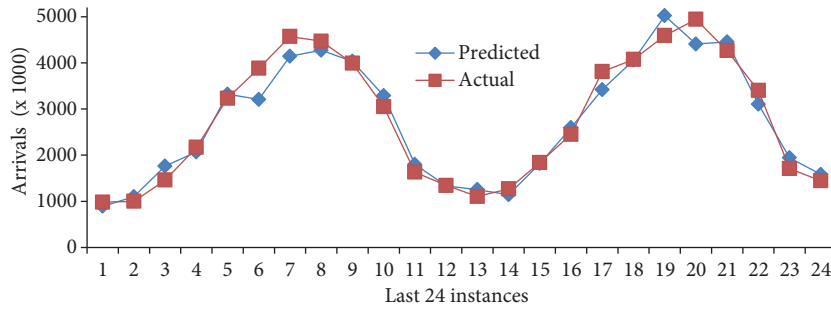


Figure 3. Comparison of actual values with the results obtained from SVR model ID 13.

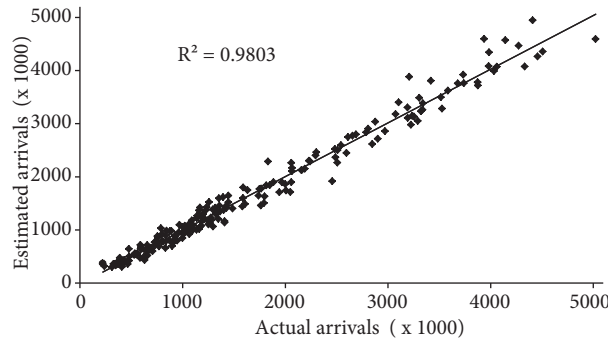


Figure 4. Performance of SVR model ID 13.

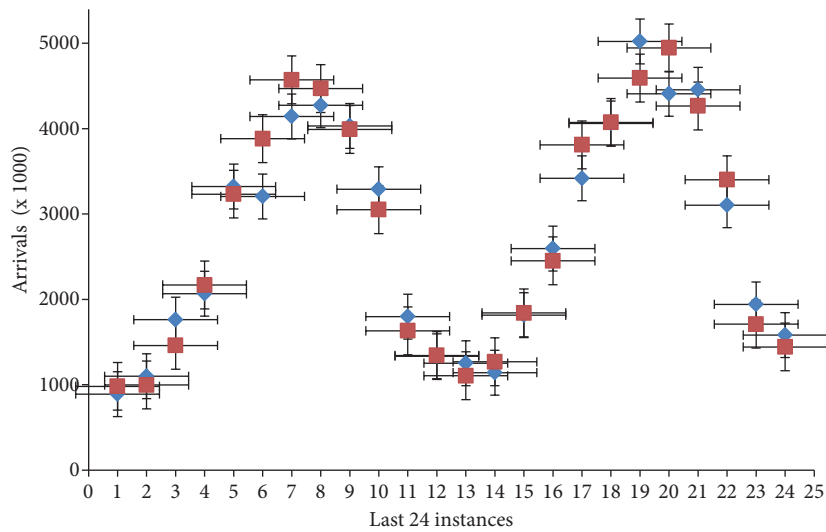


Figure 5. Error bars for SVR model ID 13.

3.3. Comparison of the models

In the second step of our study we compared the forecasting performances of the best models in the SVR, MLP regression, and MLR categories.

SVM differs from the other data mining methods as a neural network with its generalization technique. Instead of minimizing calculated training errors, SVM attempts to minimize the generalized error bound to achieve generalized performance [42,43]. While SVR and MLR deterministically converge to the same solution for a given data set, MLP has a stochastic training process since it starts with randomly assigned ANN connection parameters [44,45].

For this reason, to compare the performance of the best models in the categories of MLP, SVR, and MLR, simulations were repeated 10 times in order to obtain statistically significant comparisons. In each new simulation, the data set was randomized again. Because we used 10-fold cross-validation in each simulation, 10 different independent simulations produced 100 reliable and realistic error estimates. In other words, for statistical confidence, the training and testing process was repeated 100 times with the data sets randomly split by 10-fold cross-validation. In all cases, we reported the averages of the testing errors and next to them the standard deviation values in parenthesis (Table 3). Moreover, the statistical significance among the best models of each category according to their performance was measured by corrected two-tailed paired t-test at the $\alpha = 0.05$ significance level to identify if there is significant difference between the models. The averages of 100 error estimates across all folds (10-fold) and runs (10 independent simulations) were reported with standard deviation values.

Table 3. Overall performances of all the methods.

Model	Rank	Correlation coefficient*	Relative absolute error*	Root relative squared error*
MLR	3	0.98 (0.02)	16.62% (5.16)	20.62% (13.63)
MLP Regression	2	0.99 (0.01)	14.96% (4.03)	16.16% (5.18)
SVR	1	0.99 (0.01)	12.96% (3.28)	14.51% (3.85)

* The results are given in the form means \pm standard deviation.

The results obtained in our research indicate that SVR provides more accurate forecasts than MLP regression and MLR methods. According to the simulation results, SVR models with PUK kernel have satisfactory precision, which is even better than those of MLP regression and classical linear regression models. Comparison of the 3 best models is shown in Figure 6 and Tables 3 and 4, where it can be seen that the SVR model is the best and the MLP regression model is better than the MLR model. The WEKA experimenter module was reported as a result of corrected paired two-tailed t-test at $\alpha = 0.05$ significance level; there was a significant difference between the SVR and MLP regression models and also there was a significant difference between the MLP regression and MLR models in terms of the RAE measure. We ranked the models in Table 3 by using corrected paired two-tailed t-test in terms of the RAE results.

3.4. Discussion

In this study, the simulation results support the discussions in the “Introduction” section of this paper. As seen in Table 3, apparently SVR has better performance than the MLR and the MLP regression.

As shown in Figure 6, the data for the model of tourism demand forecasting for Turkey only takes on a monotonously increasing or decreasing change. Usually, the bigger the fluctuation of data is, the bigger the error is. SVR uses the principle of structural risk minimization approach in place of experiential risk minimization,

which makes it an excellent generalization method in the case of a small sample. That is why SVR has an ability to generalize even in the wake of fluctuation of data [38]. Thus, the forecasting error of SVR is also small under circumstances in which changes in the data take on great fluctuation. The ANN, which implements the principle of empirical risk minimization [38], needs a large amount of training data, so the forecasting performance of ANN is less excellent than that of SVR in the case of a small sample [46].

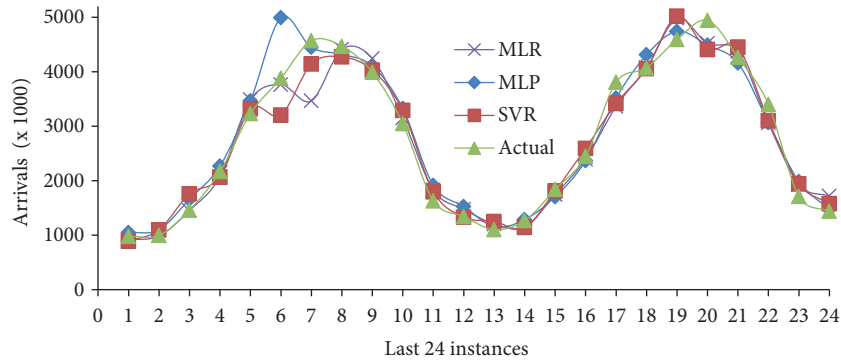


Figure 6. Comparison of all methods.

Table 4. Results of the best MLR, MLP, and SVR models for the last 12 instances.

	Predicted		
Actual	MLR	MLP	SVR
1104754	1208568	1156954	1252545
1268440	1147012	1287479	1141685
1841154	1747768	1707569	1815986
2451031	2397896	2368385	2595253
3810236	3365125	3515607	3418619
4073906	4075008	4313579	4060079
4593511	4981012	4744824	5022788
4945999	4520235	4490523	4409580
4266133	4373720	4161433	4456448
3402460	3061817	3067772	3104351
1709479	1928022	1983723	1942104
1442995	1718883	1477034	1582282

Because the superior performance on the training stage does not always guarantee generalization achievement, the performance of the developed models was measured using test data, which is out-of-sample data. The accuracy of the forecasting model must be evaluated with unbiased performance comparisons. One possible approach for evaluating the forecasting performance is to investigate whether traditional error measures such as correlation coefficient (R^2) and error bars between the actual out-of-sample returns and their predicted values are small or highly correlated, respectively. The empirical results show that SVR can accurately estimate the tourism demand for Turkey because of the high correlation (R^2) and low error bars as seen in Figures 4 and 5. This is due to the fact that the correlation (R^2) and error bars of these models indicate higher positive relationship between the actual and predicted values of the tourism demand for Turkey. The findings strongly support the nonlinearity relationship between the past time series [47,48].

The study shows that SVR provides a promising alternative to the ANN for the tourism demand forecasting. As demonstrated in the simulations, SVR forecasts considerably better than the ANN in the time series prediction. The superior performance of SVR over ANN is due to the following reasons:

- 1) SVR realizes the structural risk minimization principle, which minimizes an upper bound of the generalization error rather than minimizing the training error. This eventually leads to better generalization than the ANN, which implements the empirical risk minimization principle [42].
- 2) Since the gradient steepest descent algorithm utilized in the ANN method updates the weights in such a way that the summed square error is minimized along the steepest slope of the error surface, the ANN may not converge to global solutions. Furthermore, a global solution is not guaranteed since the algorithm can become stuck in the local minima. On the contrary, the training SVR is equivalent to solving a linearly constrained quadratic programming and the solution of SVR is always unique, optimal, and global [49,50].
- 3) The usage of the validation set to end the training of the ANN needs much experience and care. Although we have a validation set, it is quite problematic to guarantee that there is no overfitting in the ANN. This is a weakness of this method. Early stopping during training may not permit the network to learn the complexity of the prediction. On the other hand, stopping training too late may let the network learn the complexity too much, resulting in overfitting the training samples. Even though we have the advantage of using the validation set, it is still challenging to determine whether there is overfitting or not in the ANN.

4. Conclusion

In the study of multivariate processes, a framework is needed for describing not only the properties of the individual series, but also the possible cross-relationship among the series. The purpose of analyzing and modeling the series jointly is to understand the dynamic relationships over time between the series and to improve the accuracy of forecasts of individual series by utilizing the additional information available from the related series in the forecasts for each series [51].

This study presents a multivariate time series forecasting to predict the tourism demand for Turkey by employing MLR, SVR, and MLP regression methods. The real data sets with respect to Turkey and its top ranked 26 tourism client countries are used to compare the performance of those methods and to determine their achievement on forecasting tourism demand for Turkey. Comparison of the simulation results among the MLR, SVR, and MLP regression demonstrated that the SVR method has the best forecasting accuracy. Simulation results showed that the SVR model can produce lower prediction error and higher prediction accuracy and outperforms the MLR and MLP regression models. According to the simulations, it can be concluded that the tuned SVR method with the multivariate time series has enough satisfactory performance to forecast the tourism demand for Turkey; however, we still need numerous simulations to evaluate and determine the most suitable SVR model.

This paper compared the performance of different data mining methods in forecasting tourism demand for Turkey. Unfortunately, there is no certain or systematic method to select the appropriate model. Tourism planners and managers face uncertainty over the short and long run in the future demand for tourism goods and services and developing accurate forecasting tools would be of great help in planning and management. Our results showed that among the methods mentioned above, SVR has better performance and can be employed in multivariate time series forecasting.

References

- [1] Archer B. Demand forecasting and estimation. In: Ritchie JRB, Charles RG, editors. *Travel, Tourism, and Hospitality Research - A Handbook for Managers and Researchers*. New York, NY, USA: John Wiley & Sons, 1987. pp. 77-85.
- [2] Petropoulos C, Nikolopoulos K, Patelis VA. A technical analysis approach to tourism demand forecasting. *Appl Econ Lett* 2005; 12: 327-333.
- [3] Pai PF, Hong WC. An improved neural network model in forecasting tourist arrivals. *Ann Tourism Res* 2005; 32: 1138-1141.
- [4] Aslan A, Kaplan M, Kula F. International tourism demand for Turkey: a dynamic panel data approach. *Research Journal of International Studies* 2009; 9: 65-73.
- [5] Turkish Tourism Industry Report. Investment Support and Promotion Agency of Turkey. Prime Ministry of the Republic of Turkey, 2010.
- [6] World Tourism Organization. UNWTO Tourism Highlights. UNWTO, 2013.
- [7] Granger CWJ, Terasvirta T. *Modelling Nonlinear Economic Relationships*. Oxford, UK: Oxford University Press, 1993.
- [8] Efendigil T, Önit S, Kahraman C. A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: a comparative analysis. *Expert Syst Appl* 2009; 36: 6697-6707.
- [9] Gooijer JGD, Hyndman RJ. 25 years of time series forecasting. *Int J Forecasting* 2006; 22: 443-473.
- [10] Zhang GP, Patuwo EB, Hu MY. Forecasting with artificial neural networks: the state of the art. *Int J Forecasting* 1998; 14: 35-62.
- [11] Song H, Li G. Tourism demand modelling and forecasting - a review of recent research. *Tourism Manage* 2008; 29: 203-220.
- [12] Witt SF, Witt CA. Forecasting tourism demand: a review of empirical research. *Int J Forecasting* 1995; 11: 447-475.
- [13] Aslanargun A, Mammadov M, Yazici B, Yolacan S. Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. *J Stat Comput Sim* 2007; 77: 29-53.
- [14] Chen KY, Wang CH. Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Manage* 2005; 28: 215-226.
- [15] Vapnik V, Golowich S, Smola A. Support vector machine for function approximation regression estimation and signal processing. *Adv Neur In* 1996; 9: 281-287.
- [16] Hong WC. Electric load forecasting by support vector model. *Appl Math Model* 2009; 33: 2444-2454.
- [17] Lin KP, Pai PF. A fuzzy support vector regression model for business cycle predictions. *Expert Syst Appl* 2010; 37: 5430-5435.
- [18] Uysal, M. The determinants of tourism demand: a theoretical perspective. In: Debbage KG, Ioannides D, editors. *The Economic Geography of the Tourist Industry: A Supply-Side Analysis*. London, UK: Routledge, 1998. pp. 79-95.
- [19] Haykin SO. *Neural Networks: A Comprehensive Foundation*. 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1998.
- [20] Jones MT. *Artificial Intelligence: A Systems Approach*. Hingham, MA, USA: Infinity Science Press LLC, 2008.
- [21] Bishop CM. *Neural Networks for Pattern Recognition*. Oxford, UK: Clarendon press, 1995.
- [22] Vapnik V. *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.
- [23] Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2005.
- [24] Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support vector regression machines. *Adv Neural In* 1997; 9: 155-161.

- [25] Kecman V. Learning and soft computing: support vector machines, neural networks and fuzzy logic models. Cambridge, MA, USA: The MIT Press, 2001.
- [26] McNelis PD. Neural Networks in Finance: Gaining Predictive Edge in the Market. Burlington, MA, USA: Elsevier Academic Press, 2005.
- [27] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* 2009; 11: 10-18.
- [28] Choy KL, Lee WB, Lo V. Design of an intelligent supplier relationship management system: a hybrid case based neural network approach. *Expert Syst Appl* 2003; 24: 225-237.
- [29] Lenard MJ, Alam P, Madey GR. The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decision Sci* 1995; 26: 209-227.
- [30] Piramuthu S, Shaw M, Gentry J. A classification approach using multilayered neural networks. *Decis Support Syst* 1994; 11: 509-525.
- [31] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, editors. *Parallel Distributed Processing*. Cambridge, MA, USA: MIT Press, 1986. pp. 318-362.
- [32] Palmer A, Montano JJ, Sese A. Designing an artificial neural network for forecasting tourism time series. *Tourism Manage* 2006; 27: 781-790.
- [33] Platt JC. Fast training of support vector machines using sequential minimal optimization. In: Schölkopf B, Burges C, Smola A, editors. *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA, USA: MIT Press, 1998. pp. 185-208.
- [34] Schölkopf B, Smola AJ. *Learning with Kernels - Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.
- [35] Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK. Improvements to SMO algorithm for SVM regression. *IEEE Neural Networ* 2000; 11: 1188-1193.
- [36] Duan K, Keerthi SS, Poo AN. Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing* 2003; 51: 41-59.
- [37] Keerthi SS. Efficient tuning of SVM hyper-parameters using radius/margin bound and iterative algorithms. *IEEE Neural Networ* 2002; 13: 1225-1229.
- [38] Fei SW, Liu CL, Miao YB. Support vector machine with genetic algorithm for forecasting of key-gas ratios in oil-immersed transformer. *Expert Syst Appl* 2009; 36: 6326-6331.
- [39] Cristianini N, Shawe-Taylor J. *An introduction to support vector machines and other kernel-based learning methods*. New York, NY, USA: Cambridge University Press, 2000.
- [40] Gunn SR. *Support Vector Machines for Classification and Regression*. Southampton, UK: University of Southampton, 1997.
- [41] Vapnik V. *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.
- [42] Basak D, Pal S, Patranabis DC. Support vector regression. *Neural Information Processing – Letters and Reviews* 2007; 11, 203-224.
- [43] Deng S, Liu DM, Hsieh SL. Applying machine learning methods to the shooting accuracy prediction: a case study of T-75 pistol shooting. *Information Technology Journal* 2011; 10: 1508-1517.
- [44] Campbell C, Cristianini N. Simple training algorithms for support vector machines. Technical Report CIG-TR-KA, University of Bristol, Engineering Mathematics, Computational Intelligence Group, 1999.
- [45] Brown MPS, Grundy WN, Lin D, Cristianini N, Sugnet C, Ares M, Haussler D. Support Vector Machine Classification of Microarray Gene Expression Data. Technical Report UCSC-CRL-99-09, University of California, Santa Cruz, CA, USA, 1999.

- [46] Fei SW, Wang MJ, Miao YB, Tu J, Liu CL. Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil. *Energ Convers Manage* 2009; 50: 1604-1609.
- [47] Pesaran MH, Timmermann A. Predictability of stock returns: robustness and economic significance. *J Financ* 1995; 50: 1201-1227.
- [48] Thawornwong S, Enke D. The adaptive selection of financial and economic variables for use with artificial neural networks. *Neurocomputing* 2004; 56: 205-232.
- [49] Cheng J, Qian JS, Niu GD, Guo YN. Gas concentration forecasting based on support vector regression in correlation space via KPCA. In: Yu W, He H, Zhang N, editors. *Advances in Neural Networks – ISNN, 2009*. Berlin, Germany: Springer-Verlag, 2009, pp. 937-945.
- [50] Li D, Mersereau RM, Simske S. Blind image deconvolution through support vector regression. *IEEE Neural Networ* 2007; 18: 931-935.
- [51] Reinsel GC. *Elements of Multivariate Time Series Analysis*. 2nd ed. New York, NY, USA: Springer-Verlag, 1997.