International Journal of Financial, Accounting, and Management (IJFAM)
ISSN: 2656-3355, Vol 6, No 4, 2025, 509-524 https://doi.org/10.35912/ijfam.v6i4.2492

Optimum estimation and forecasting of gasoline consumption in Iran's national oil refining and distribution company

Afshar Bazyar[1*], Morteza Abbasi[2]
Malek Ashtar University of Technology, Iran[1&2]
bazyar@mut.ac.ir[1], mabbasi@mut.ac.ir[2]

Article History
Received on 30 October 2024
1st Revision on 17 November 2024
2nd Revision on 26 November 2024
3rd Revision on 10 December 2024
4th Revision on 17 December 2024
5th Revision on 28 December 2024
6th Revision on 7 January 2025
7th Revision on 26 January 2025
8th Revision on 8 February 2025
Accepted on 17 February 2025

Abstract
Purpose: This paper presents an accurate estimation and forecasting of gasoline consumption. This is vital for the policy and decision-making process in the energy sector.
Method: A hybrid data-driven model based on Artificial Neural Network (ANN) and an autoregressive integrated moving average (ARIMA) approach was developed for optimum estimation and forecasting of gasoline consumption. The proposed hybrid ARIMA-ANN approach considers six lagged variables and one forecasted value provided by the ARIMA process. The ANN trains and tests data with a multi-layer perceptron (MLP) approach, which has the lowest Mean Absolute Percentage Error (MAPE). To show the applicability and superiority of the proposed hybrid approach, daily available data were collected for 7 years (2015–2021) in Iran.
Results: The acquired results show a high accuracy of about 94.27% using the proposed hybrid ARIMA-ANN method. The results of the proposed model are compared with respect to the regression models and the ARIMA process.
Conclusions: Analyzing consumption patterns can provide insights into consumer behavior, enabling NIORDC to tailor its services and marketing strategies more effectively.
Limitations: Eliminating subsidies from gasoline prices has led to the appearance of noisy data in gasoline consumption in Iran's National Oil Refining and Distribution Company.
Contribution: The outcome of this paper justifies the capability of the proposed hybrid ARIMA-ANN approach in accurate forecasting of gasoline consumption.

Keywords: *Gasoline consumption; Artificial Neural Networks; ARIMA; Forecasting; Multi Layer Perceptron*

How to Cite: Bazyar, A., & Abbasi, M. (2025). Optimum estimation and forecasting of gasoline consumption in Iran's national oil refining and distribution company. *International Journal of Financial, Accounting, and Management,* 6(4), 509-524.

1. Introduction

Iran's National Oil Refining and Distribution Company (NIORDC) is a state-owned entity responsible for managing the refining, production, distribution, and retailing of petroleum products in Iran. Operating under the Ministry of Petroleum, NIORDC plays a critical role in ensuring the country's energy security, managing its domestic oil supply chain, and regulating the distribution of refined petroleum products, including gasoline, diesel, and liquefied petroleum gas (LPG). One of the key roles and responsibilities of NIORDC is gasoline and diesel Supply. A significant portion of Iran's refined products consists of gasoline and diesel, which are crucial for the country's transport sector. NIORDC ensures that there is enough supply of these fuels to meet the needs of the population, especially given the large number of vehicles in the country. Moreover, through a vast network of gas stations across Iran, NIORDC provides retail fuel services to consumers. The company directly manages this network or works with third-party operators, ensuring that fuel is available to the public at subsidized prices (especially in the case of gasoline). Therefore NIORDC is responsible for maintaining and upgrading the infrastructure needed to support the oil refining and distribution processes. This includes refineries, storage facilities, pipelines, and fuel depots, as well as developing new projects to increase refining capacity or improve fuel quality. On the other hand, Iran's refining capacity is not always sufficient to meet all domestic fuel needs. In such cases, NIORDC imports refined petroleum products from abroad. Conversely, the company also participates in the export of surplus refined products, though international sanctions have often complicated these efforts.

Time series forecasting is an interesting research area that has attracted many practitioners and researchers over the past several decades. In time series forecasting approaches, historical observations of the same variable are analyzed to extract the most appropriate model describing the relationship between the current data and the past observed data. The time series modeling approach is used when there is no exhaustive model linking the prediction variable to other explanatory variables with high accuracy or when there is little knowledge about the underlying data generating the process (Khashei & Bijari, 2011). The autoregressive integrated moving average (ARIMA) is one of the prevalent linear models used in forecasting energy, engineering, exchange rate and stock problems. Moving average and exponential smoothing are other tools used in linear forecasting.

ARIMA models are composed of the pure autoregressive (AR), the pure moving average (MA), and a combination of the AR and MA (ARMA). ARIMA models are able to represent different types of time series, i.e., AR, MA, and ARMA series. Nevertheless, since the complex real-world problems usually have a nonlinear structure, the pre-assumed linear correlation structure among the time series values is their major drawback in real-world applications (Ghahnavieh, 2019).

The use of the ANN has proved its efficiency as an estimation tool for predicting factors through other input parameters that have no specified relationship. Some examples of this work are provided in Azadeh, Asadzadeh, and Ghanbari (2010) and Tarafdar and Ghadimi (2014). In addition the capabilities of ANN methods help us to gain more reliable results (Azadeh, Babazadeh, & Asadzadeh, 2013; Fathi, Eftekhari Yazdi, & Adamian, 2020). In this study, we have introduced seven parameters including six lagged variables and one forecasted value of gasoline consumption specified by The ARIMA model. The output is the daily gasoline consumption. We applied these input parameters in the framework of the ANN, and the data have been tested and trained by Multi Layer Perceptron (MLP). Comparison of the acquired results of this study with the regression prediction models and ARIMA model shows a considerable improvement in the error amount and accuracy of the prediction. As an instance case study, we collected daily data for 7 years (2015–2021) in Iran. In this work, we provide a prediction with a acceptable amount of error that is obtained with regard to more available input data. The aim of the proposed hybrid ARIMA-ANN model is to reduce the risk of using an inappropriate model by combining several models to decrease the risk of failure and obtain results that are more accurate. The main difference of the proposed approach with respect to those existing in the literature (Díaz-Robles et al., 2008) is that this paper uses autocorrelation function (ACF) and the partial the autocorrelation function (PACF) for suitable recognition of the inputs of the ANN model.

2. Literature review

2.1 Forecasting gasoline consumption

In the National Iranian Oil Refining and Distribution Company (NIORDC), gasoline consumption is a critical focus due to its significant economic, environmental, and operational implications. NIORDC, which is responsible for refining, distributing, and managing Iran's petroleum products, views gasoline consumption as a core area affecting both the country's energy policy and economic stability.

Here is why gasoline consumption is very important in NIORDC’s operations:

Economic Impact: Gasoline is a major revenue generator for Iran's energy sector. Optimizing consumption directly affects the cost-efficiency of domestic and export fuel markets, thereby impacting national revenue streams. By managing consumption, NIORDC helps reduce dependence on imports, especially during times of high domestic demand, which contributes to energy security and economic resilience.

1. Energy Sustainability: As gasoline demand grows, NIORDC must balance production and consumption with sustainability goals. Managing gasoline usage is critical for ensuring long-term energy availability, reducing pressure on refinery capacity, and helping meet government-set targets to decrease fossil fuel dependence.
2. Environmental Concerns: Iran faces environmental challenges associated with high fossil fuel consumption. By encouraging efficient gasoline use, NIORDC can help reduce greenhouse gas emissions, improve urban air quality, and align with environmental regulations. This is especially important as Iran addresses pollution and other environmental issues in major cities.
3. Technology and Innovation in Distribution: Controlling gasoline consumption often requires technological investments in refining and distribution. NIORDC invests in technologies that monitor and enhance fuel efficiency, aiming to streamline distribution channels and minimize fuel loss, which also impacts environmental sustainability.
4. Energy Security and Policy: Gasoline consumption trends affect Iran’s energy security policies. NIORDC analyzes consumption data to help develop policies that align with national goals, such as reducing energy subsidies or promoting alternative fuels. These policies are essential for ensuring the country’s energy independence and reducing its vulnerability to global fuel market fluctuations.
5. Public Awareness and Efficiency Programs: The NIORDC runs programs to encourage the public to use gasoline more efficiently, which helps reduce overall demand. This includes public campaigns and sometimes implementing fuel rationing, especially during periods of economic strain, to stabilize consumption levels.

By managing gasoline consumption, NIORDC not only supports Iran’s economic health and energy independence but also advances environmental goals critical for the country's sustainable development.

Mistakes in forecasting gasoline consumption can have significant consequences for energy management, economic stability, and environmental sustainability. Here are key reasons why accuracy in gasoline consumption forecasting is critical and why mistakes can be costly economic and Budgetary Impact such as: supply and demand mismatch, shortages and budgetary strain. Inaccurate forecasts can lead to supply-demand mismatches, with either shortages or surpluses of gasoline. Shortages can disrupt transportation and economic activities, while surpluses can lead to costly storage issues or losses. Mistakes in forecasting can cause budget allocation issues, as the costs of emergency imports or underutilized refineries may put a strain on government resources, impacting other areas of public spending (Mbamalu, Chike, Oguanobi, & Egbunike, 2023; Zahedi, Abbasi, & Khanachah, 2020).

Mistakes in forecasting gasoline consumption Impacts on Import and Export Strategies. If demand is underestimated, it may require importing additional fuel, often at higher prices. This reliance can be financially burdensome and reduce energy security. Moreover excess gasoline that could have been exported may go unused or sold at lower rates, reducing potential revenue.

One of the adverse effects of errors in forecasting gasoline consumption is its environmental impact. Surplus gasoline often leads to lower prices, which can encourage higher consumption, resulting in increased emissions and environmental impact. Conversely, shortage may encourage the use of less efficient or environmentally harmful fuel substitutes. Moreover, forecast errors can skew sustainability initiatives, as policies aimed at reducing consumption might be insufficient if consumption spikes unexpectedly, undermining environmental objectives.

Inaccurate forecasting can result in reduced efficiency in operations and infrastructure. Mistakes can result in overuse or underuse of refining capacities. Overuse stresses the infrastructure, leading to higher maintenance costs, while underuse results in wasted operational capacity and inefficiency. Forecasting errors can lead to bottlenecks or misdistribution across regions, causing regional shortages or gluts that disrupt efficient fuel distribution.

Due to errors in forecasting gasoline consumption or incorrect estimations, strategic planning and policymaking cannot be effectively performed. Forecasting mistakes impact energy security by affecting inventory levels and emergency reserves. Misjudging the balance between supply and demand could leave reserves depleted, especially in emergencies. Policy measures, such as pricing adjustments, subsidies, or conservation programs, rely on accurate forecasts. Forecast errors can lead to misguided policies that fail to effectively address consumption patterns or market needs.

Erroneous diesel consumption forecasts can weaken national trust and diminish consumer confidence. Shortages resulting from forecast errors can inconvenience consumers, leading to long wait times at fuel stations and potential public dissatisfaction. Repeated forecasting errors may erode public trust in national energy management, making it more challenging to implement conservation programs or energy policies (Forozandeh, 2021).

Mistakes in forecasting gasoline consumption have ripple effects that reach beyond immediate supply and demand issues, impacting economic stability, environmental goals, infrastructure health, and even public trust. Ensuring accurate forecasts through improved modeling, data quality, and adaptable strategies is essential for stable, sustainable, and efficient gasoline consumption management.

Incorrect forecasting of diesel consumption can lead to disruptions, as it may result in supply shortages or surpluses, causing instability in the market. This can fuel public dissatisfaction, leading to protests or social unrest, particularly if consumers face price hikes or fuel shortages. The inability to accurately predict fuel needs can also undermine trust in the government's ability to manage resources effectively, further escalating tensions and contributing to societal disturbances. In Iran, inaccurate diesel consumption forecasts can directly contribute to fuel smuggling. Overestimating consumption may result in surplus fuel, which can then be illegally diverted. On the other hand, underestimating demand can lead to fuel shortages, prompting smuggling as people or groups seek to sell fuel at higher prices in neighboring countries. These forecasting errors create instability in the fuel supply, increasing the chances of fuel being diverted illegally, which poses a threat to both economic stability and energy security.

Forecasting gasoline consumption is a complex task that involves predicting future demand based on various factors such as historical consumption patterns, economic conditions, population growth, and external influences like fuel prices and government policies. Different forecasting methods can be employed depending on the time frame, data availability, and the level of accuracy required. Here are the most commonly used methods for forecasting gasoline consumption.

2.2. forecasting methods

2.2.1. Time Series Forecasting Methods

Time series methods involve analysing historical data on gasoline consumption to identify patterns, trends, and seasonal variations. These methods are useful when historical data is available and the future is expected to follow similar patterns to the past.

*a) **Moving Averages***

i. **Simple Moving Average (SMA)**: The simplest method, where past consumption data is averaged over a specific period (e.g., 12 months). The forecast is based on the assumption that future consumption will follow the average trend of past data.
ii. **Weighted Moving Average (WMA)**: Similar to SMA but with more weight given to recent data points. This method is useful when more recent consumption trends are more reflective of future demand.

It may suitable method in short-term forecasts, especially in stable demand environments without major disruptions.

b) Exponential Smoothing

i. **A method where more weight is given to recent data, with the smoothing constant** determining the level of emphasis on newer data. There are different variations, such as:
   a. **Single Exponential Smoothing (SES):** Suitable for data without a clear trend or seasonality.
   b. **Double Exponential Smoothing (DES):** Used when there is a trend in the data.
   c. **Triple Exponential Smoothing (TES):** Also known as **Holt-Winters method**, this method accounts for both trend and seasonality in data, making it suitable for gasoline consumption that exhibits these characteristics.

It can be suitable method in medium to short-term forecasts with moderate trend and seasonal variations.

c) Autoregressive Integrated Moving Average (ARIMA)

ARIMA models are used when the time series data shows complex patterns (e.g., trends, seasonality, and autocorrelation). ARIMA models combine autoregression (AR), differencing (I), and moving averages (MA) to make predictions based on historical data.

It may suitable method in Longer-term forecasts when the data shows patterns like seasonality, trend, and autocorrelation.

2.2.2. Causal (Econometric) Forecasting Models

Causal models look beyond historical consumption data and take into account external variables or drivers that influence gasoline consumption, such as economic indicators, fuel prices, demographic factors, or government policies.

a) Linear Regression

i. Linear regression is one of the simplest causal models. It models the relationship between gasoline consumption and one or more independent variables (e.g., GDP, population size, vehicle ownership, fuel prices). The goal is to identify how changes in these variables are likely to affect gasoline consumption.
ii. **Multiple Linear Regression**: This is an extension of simple linear regression, where multiple factors (e.g., income levels, fuel prices, vehicle fleet size) are considered simultaneously to predict gasoline demand.

It may suitable method when there is a clear relationship between gasoline consumption and specific economic or demographic factors.

b) Multivariate Regression Models

Multivariate regression models allow for the inclusion of several independent variables that can jointly affect gasoline consumption. These models may include factors such as GDP growth, fuel prices, inflation, consumer spending, and technological advancements in vehicle fuel efficiency.

It may suitable method when gasoline consumption is driven by multiple interacting factors, such as economic growth, urbanization, and policy changes.

c) System Dynamics Models

These models simulate the complex feedback systems that affect gasoline consumption. System dynamics models can take into account long-term factors like population growth, technological change, and the environmental impact of fuel consumption, which may influence government policies or consumer behavior over time.

It may be a suitable method in long-term forecasts that consider a wide range of social, economic, and environmental factors.

2.2.3. Machine Learning and Artificial Intelligence (AI) Models

In recent years, machine learning and AI techniques have gained traction in forecasting tasks. These models can capture complex, nonlinear relationships in large datasets and adapt to changing conditions. They are especially useful when dealing with large amounts of data or non-linear patterns.

a) ***Artificial Neural Networks (ANN)***

ANNs can learn from historical data and detect patterns or trends in gasoline consumption that may not be apparent using traditional methods. They work well with complex data structures and can handle large datasets, making them effective for forecasting gasoline demand over time.

It may suitable method when data is large, complex, and includes multiple variables that interact with one another (e.g., weather patterns, fuel prices, vehicle fleet characteristics).

b) ***Random Forests and Decision Trees***

Decision trees and random forests can model non-linear relationships between gasoline consumption and influencing factors. These models break down the problem into decision nodes and branches, leading to better understanding and prediction of consumption patterns.

Use Case: When gasoline consumption is influenced by various non-linear factors that require advanced classification or regression.

c) ***Support Vector Machines (SVM)***

SVM is another machine learning technique that can be used to forecast gasoline consumption, especially when the relationship between input variables and consumption is highly non-linear.

It may suitable method in when there is high complexity and the relationships between gasoline consumption and influencing factors are not easily captured by traditional models.

2.2.4. Expert Judgment and Delphi Method

In situations where data is sparse or when external factors such as political events or technological breakthroughs play a significant role, expert judgment can be an essential forecasting tool. The **Delphi Method** is a structured approach where a panel of experts provides estimates about future gasoline consumption, and through multiple rounds of questioning, their opinions are aggregated to produce a forecast.

It may suitable method when there is limited data or when the forecast is needed in the face of uncertainty, such as during policy changes or economic crises.

2.2.5. Scenario Planning

Scenario planning involves developing multiple potential future scenarios based on different assumptions about key drivers of gasoline consumption (e.g., oil prices, technological changes, policy shifts). This method is particularly useful when there is uncertainty about future conditions but provides valuable insights into a range of possible outcomes.

It may be a suitable method in long-term forecasting in uncertain environments where several external factors might significantly impact demand, such as geopolitical shifts, technological innovation, or changes in government policy.

Each of these forecasting methods has its strengths and weaknesses, and the choice of method depends on factors like the availability of historical data, the time horizon of the forecast, the complexity of the drivers of gasoline consumption, and the specific needs of the organization. In many cases, a combination of methods is used to improve accuracy and reliability. For instance, time series methods may be used for short-term forecasting, while econometric models or machine learning techniques are applied for long-term or more complex predictions that involve multiple influencing factors.

As discussed earlier, considering the conditions and considerations of forecasting gasoline consumption in Iran, as well as reviewing various forecasting methods, we conclude that ARIMA and ANN may be the best forecasting method.

3. Research Methodology

3.1. The ARIMA and ANN forecasting models

Totally, an ARIMA model is specified via three components including order of autoregressive process (*p*), order of moving average process (*q*), and order of differencing (*d*). In an ARIMA (p, d, q) model, the future value of a variable is assumed to be a linear function of several past observations and random errors. That is, the underlying process that generates the time series with the mean μ has the form (Chike et al 2023):

$$\phi(B)\nabla^d(y_t-\mu)=\theta(B)a_t \qquad (1)$$

Where, y_t and a_t are the actual values and random error at time period *t*, respectively.

$$\phi(B)=1-\sum_{i=1}^{p}\varphi_i B^i \qquad (2)$$

$$\theta(B)=1-\sum_{j=1}^{q}\theta_j B^j \qquad (3)$$

$\phi(B)$ and $\theta(B)$ are polynomials in *B* of degree *p* and *q*, φ_i (i=1,2,…, p) and θ_j (j=1,2, …, q) are model parameters, $\nabla=(1-B)$, *B* is the backward shift operator, *p* and *q* are integers and often referred to as orders of the model, and *d* is an integer and often referred to as order of differencing. Random errors, ε_t, are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2.

The Viana, Oliveira, and Rocha (2024) methodology encompasses three iterative steps including model identification, parameter estimation, and diagnostic checking. They used ACF and PACF of the sample data as the basic tools to identify the order of the ARIMA model. We also use these functions to identify the preliminary components of the considered time series and then the most appropriate components are specified using suitable diagnostic statistical test. Recently other approaches based on intelligent paradigms, such as neural networks (Melina et al., 2024), genetic algorithms (Ong, Huang, & Tzeng, 2005) or fuzzy systems (Fan et al., 2021) have been presented to improve the accuracy of order selection of ARIMA models.

In the identification step, data transformation is often required to make the time series stationary. Stationary is a necessary condition in building an ARIMA model used for forecasting. In other words, the model estimation step is performed only after the stationary of a time series is confirmed. A stationary time series is characterized by statistical characteristics such as the mean and the autocorrelation structure being constant over time. When the observed time series show trend and heteroscedasticity, differencing and power transformation are applied to the data to remove the trend and to stabilize the variance before an ARIMA model can be fitted. Once a tentative model is identified, estimation of the model's parameters is straightforward. The parameters are estimated such that an overall measure of errors is minimized. This can be accomplished using a nonlinear optimization procedure. The last step in model building is the diagnostic checking of model adequacy. This is basically to check if the model assumptions about the errors, ε_t, are satisfied. Several diagnostic statistics and plots of the residuals can be used to examine the goodness of fit of the tentatively entertained model to the historical data. If the model is not adequate, a new tentative model should be identified which is followed by the steps of parameter estimation and model verification. Diagnostic information may suggest alternative model(s). This three-step model building process is typically repeated several times until a satisfactory model is finally selected. The final selected model can then be used for prediction purposes.

ANNs are composed of attributes that lead to perfect solutions in applications where we need to learn a linear or nonlinear mapping. Some of these attributes are: learning ability, generalization, parallel processing and error endurance. These attributes would cause the ANNs solve complex problem methods precisely and flexibly (Ho et al., 2021). ANNs consists of an inter-connection of a number of neurons. There are many varieties of connections under study, however here we will discuss only one type of network which is called the MLP. In this network the data flows forward to the output continuously without any feedback. The MLP uses a supervised learning technique called back propagation for training the network (Austina, Senthilkumar, & Kanthavelkumaran, 2013). Figure 1 shows a typical three-layer feed forward model used for forecasting purposes. The input nodes are the previous lagged observations while the output provides the forecast for the future value. Hidden nodes with appropriate nonlinear transfer functions are used to process the information received by the input nodes. The model can be written as:

$$y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j f\left(\sum_{i=1}^{m} \beta_{ij} y_{t-1} + \beta_{0j} \right) + \varepsilon_t \qquad (4)$$

Where m is the number of input nodes, n is the number of hidden nodes, f is a sigmoid transfer function such as the logistic:} is a vector of weights from n …, 0, 1, = j, $_j\alpha$. { $f(x) = \dfrac{1}{1+\exp(-x)}$ the hidden to output nodes and $\{\beta_{ij},\ i = 1,\ 2,\ \ldots,\ m;\ j = 0,\ 1,\ \ldots,\ n\}$ are weights from the input to hidden nodes. α_0 and β_{0j} are weights of arcs leading from the bias terms which have values always equal to 1. Note that Eq. (4) indicates a linear transfer function employed in the output node as desired for forecasting problems. The MLP's most popular learning rule is the error back propagation algorithm. Back Propagation learning is a kind of supervised learning introduced by Werbos (1974) and later developed by Rumelhart and McClelland (1986). At the beginning of the learning stage, all the weights in the network are initialized to small random values. The algorithm uses a learning set, which consists of input–desired output pattern pairs. Each input–output pair is obtained by the offline processing of historical data. These pairs are used to adjust the weights in the network to minimize the sum squared error (SSE) which measures the difference between the real and the desired values overall output neurons and all learning patterns.

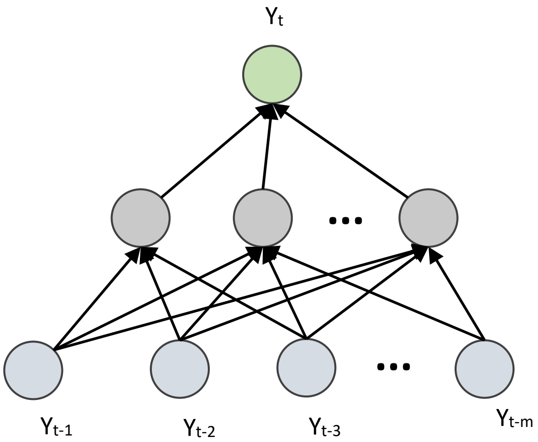

Figure 1. A three-layer MLP network to predict gasoline consumption

After computing SSE, the back propagation step computes the corrections to be applied to the weights. Most of the suggested models use MLP networks (Al Mamun et al., 2020; Khashei & Bijari, 2011). The attraction of MLP has been explained by the ability of the network to learn complex relationships between input and output patterns, which would be difficult to model with conventional algorithmic methods.

3.2. The proposed hybrid ARIMA-ANN method

The hybrid models for forecasting time series often decompose a time series into its linear and nonlinear form (Chike et al, 2023). Khashei and Bijari (2011) and Wang, Zou, Su, Li, and Chaudhry (2013) investigated the effectiveness of hybrid ARIMA-ANN methods in forecasting respect to existing forecasting methods. In hybrid models, a time series can be considered to be composed of a linear autocorrelation structure and a nonlinear component. In the present work, we consider the gasoline consumption time series as function of a linear and a nonlinear component. The amount of gasoline consumption (y_t) is a function of linear component (L_t) and nonlinear component (N_t):

$$y_t = f\,(L_t, N_t) \qquad (5)$$

In order to estimate L_t and N_t, we use the forecasted values predicted by ARIMA process. Indeed, first the valid components are chosen from the component list specified by ACF and PACF. Second, the estimated model is constructed via the valid components and used for forecasting the considered stationary time series. This phase constructs the linear component of the proposed hybrid ARIMA-ANN approach. Then, the forecasted values and specified valid components by ARIMA process are utilized to be used as input parameters of the ANN model. This phase constructs nonlinear component of the proposed hybrid ARIMA-ANN approach.

In this paper, data is collected for a robust period and is further divided to train and test groups. Train data is used to train the MLP models. Test data is used to be compared with actual data (Validation). Moreover, the best fitted MLP is identified by the lowest MAPE. In addition, the selected MLP is compared with different regression's models. Figure 2 presents the overall description of the model. Figure 3 presents the ANN pictorial of the proposed model.

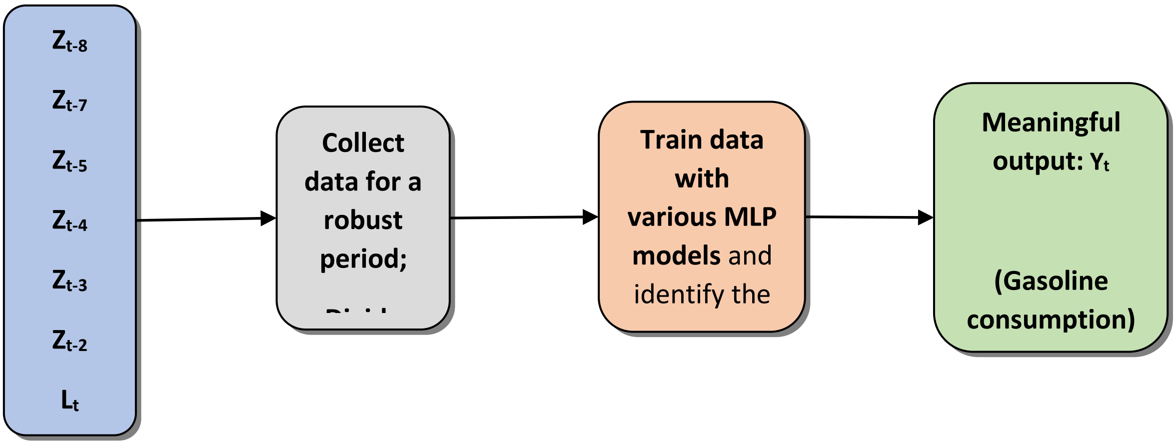

Figure 2. Description of the ANN model used to predict gasoline consumption

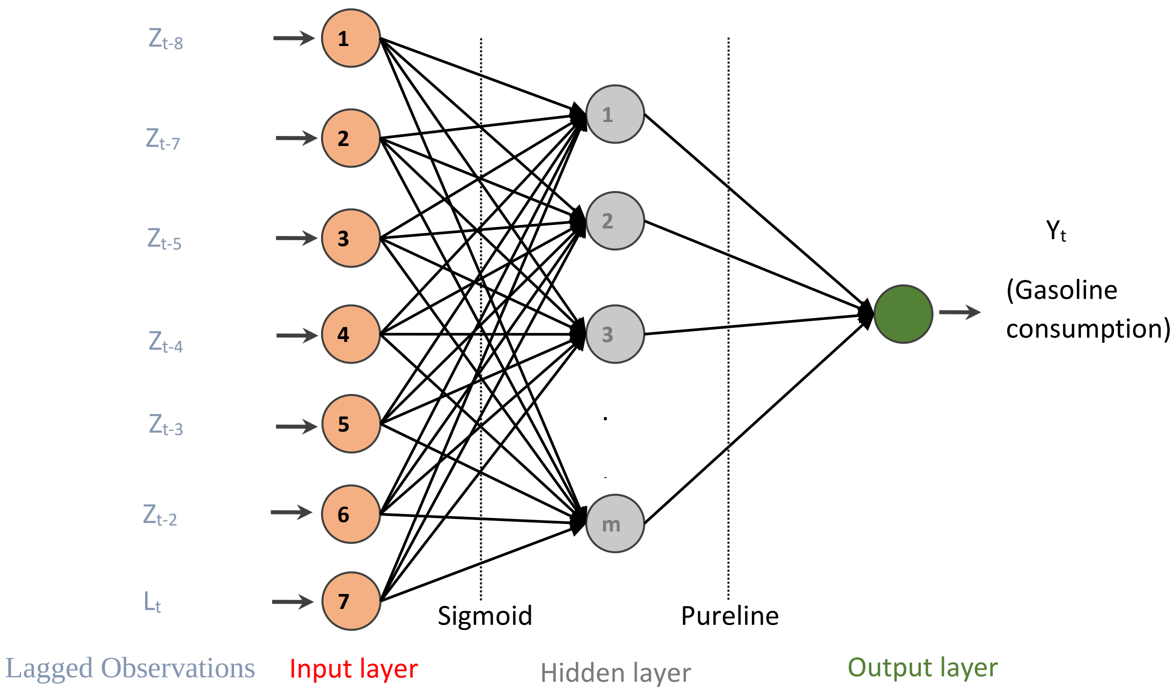

Figure 3. The integrated ANN-MLP model to predict gasoline consumption

Where Z_{t-i} is the lagged observations representing the valid component specified by ARIMA process, and L_t represents the forecasted values of gasoline consumption provided by ARIMA process. The process of extracting the most suitable lagged observations would be described in Section 3.
Among different error estimation methods we use the MAPE method to assess the performance of employed forecasting models. The MAPE calculation is as follows:

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{x_t - x'}{x_t} \right|}{n} \quad (6)$$

4. Results and discussions

The proposed model was applied in Iran. Data for these parameters are provided from Institute for International Energy Studies (IIES) in Iran and Energy Information Administration (EIA) website (http://www.eia.doe.gov/emeu/international/contents). Before going ahead, we shortly explain the most important motivations to forecast gasoline consumption in Iran.

According to “Iranian targeted subsidy plan”, The Iranian government presented an energy price reform in 2008. The major aim of the policy was to slow down the increasing trend of energy consumption in Iran by removing the energy subsidies. According to the plan, all energy prices were to increase by 20 percent annually. In 2006, daily gasoline consumption stood at 74 million liters and the country paid $5 billion for gasoline imports. The overall consumption of gasoline after the reform decreased from about 65 million liters per day to about 54 million liters per day. Therefore, accurate forecasting of gasoline consumption leads to creating insights in planning for imports, price reform, and etc.

The required data were collected daily for seven years from 2015 to 2021. We divided the dataset into two groups: the training subset and the test subset. For the training subset related data of 1851 periods (days) were considered and used for learning the model and for the test subset relevant data of 20 days were used to test the capability of the model.

4.1. The best structure of the proposed hybrid ARIMA-ANN model

Several MLP networks are generated and tested. The transfer function for the first layer and all hidden layers are sigmoid and for last one is linear. Back propagation algorithm is used to adjust the learning procedure.

The results of the ten best models and their errors are shown in Table 1. The acquired errors in the last row of Table 1 are derived for test data. According to Table 1, the best ANN structure including six neurons in the first hidden layer and four neurons in the second hidden layer is selected. This ANN structure is trained with back propagation (BP) learning algorithm. Figure 4 illustrates the actual values against forecasted values by hybrid ARIMA-ANN approach.

Table 1. Different MLP specifications and MAPE results

Model number	1	2	3	4	5	6	7	8	9	10
Number of neurons in first hidden layer	6	6	4	7	7	7	8	8	9	9
Number of neurons in second hidden layer	2	4	2	1	2	3	1	2	1	2
Learning method	BP	BP	BP	BP	BP	BP	BP	BP	BP	BP
Relative error (MAPE %)	6.74	5.73	6.33	6.32	6.07	6.30	5.76	6.6	6.07	6.22

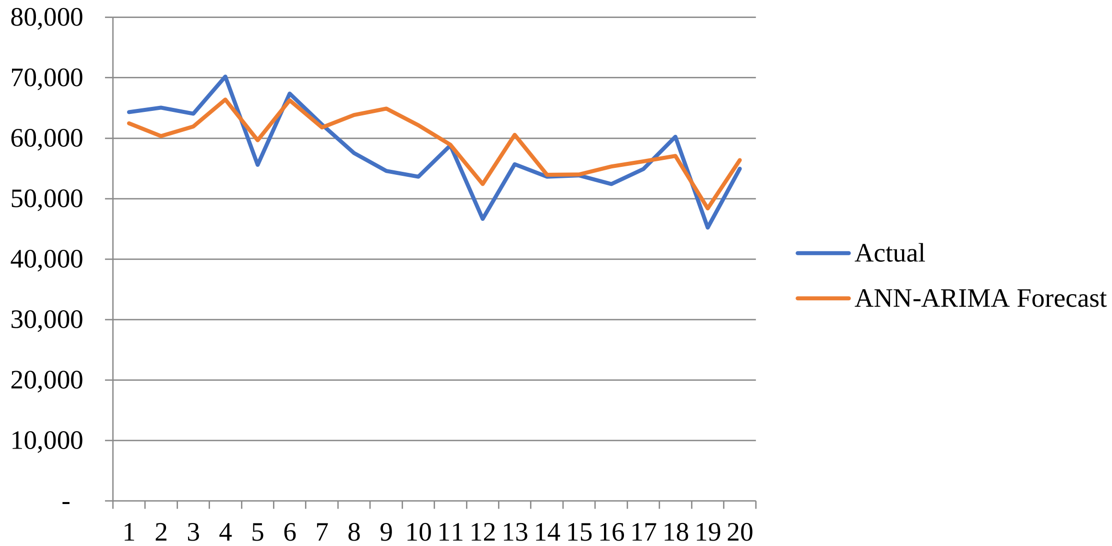

Figure 4. Actual vs. hybrid ARIMA-ANN Forecast

4.2. Implementing ARIMA process

To implement ARIMA process, we first specify the AR and MA components using autocorrelation function (ACF) and partial autocorrelation function (PACF). Variation of the considered time series has been shown in Figure 5. Stationary test on the considered time series, which is performed by diagnostic statistical test, confirms that the time series is stationary.

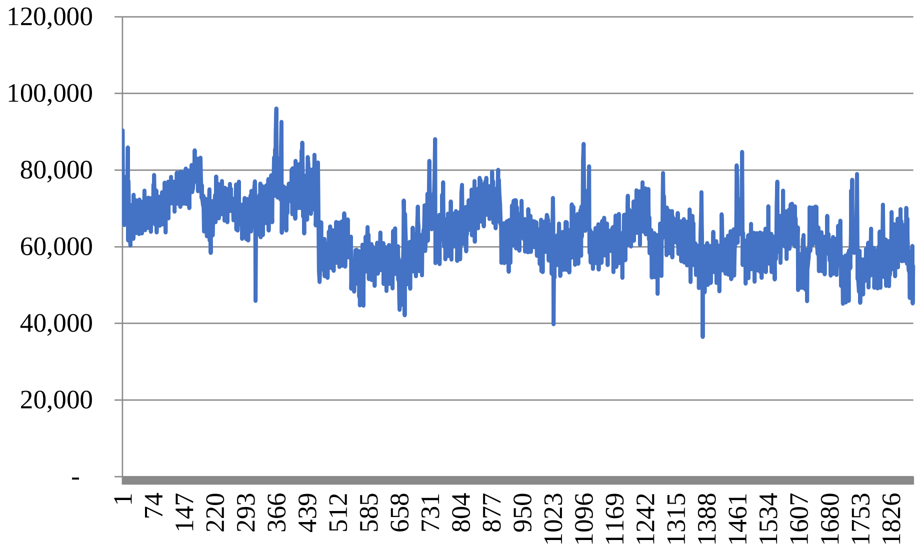

Figure 5. Variation of gasoline consumption

The ACF and PACF have been shown in figures 6 and 7 for 100 lagged observations.

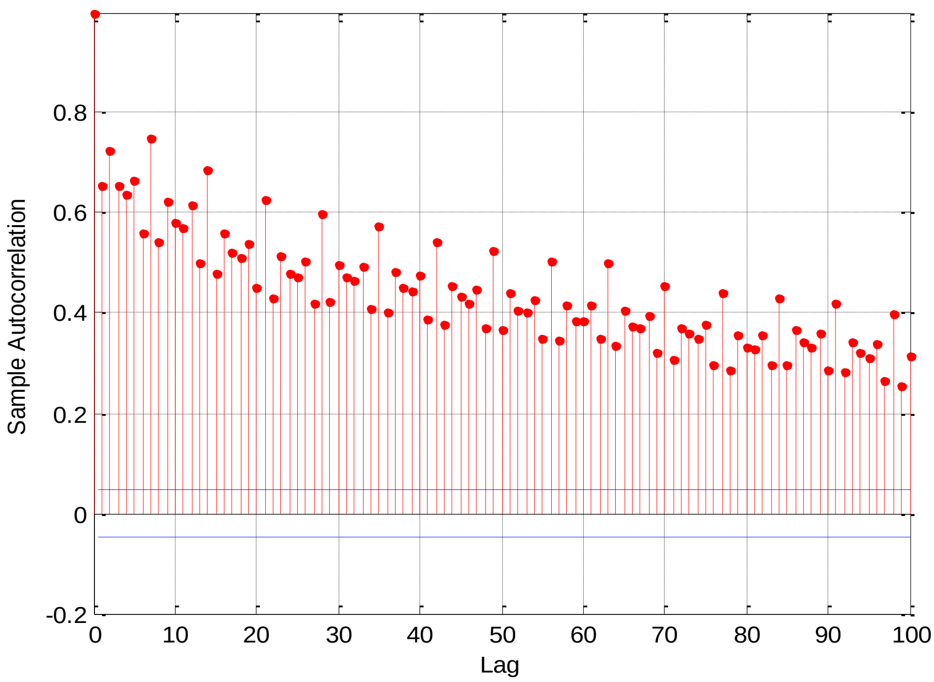

Figure 6. Autocorrelation function

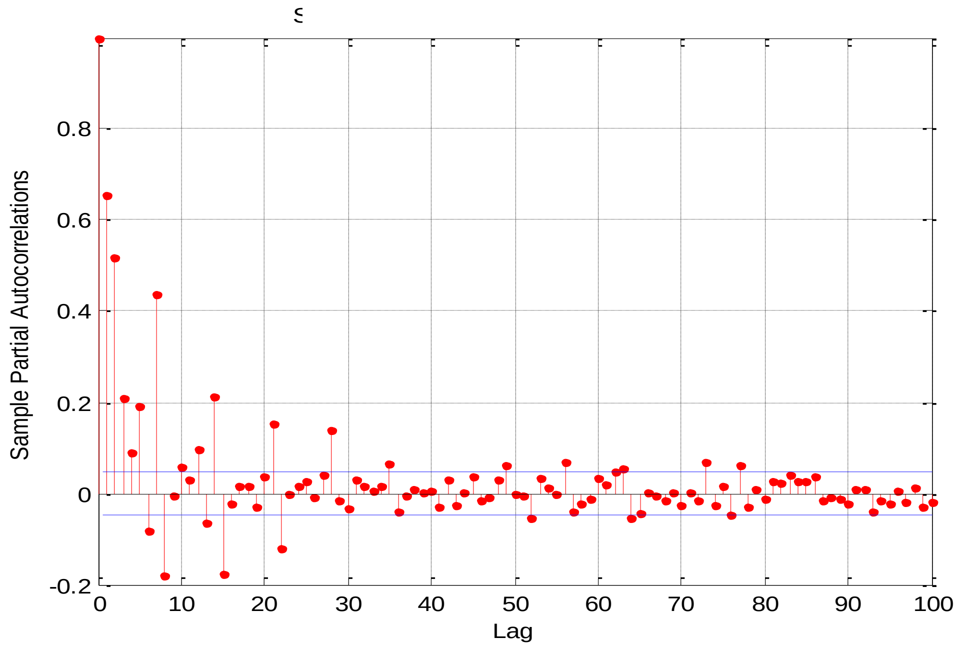

Figure 7. Partial autocorrelation function

From the ACF and PACF illustrations, the following factors are extracted to be considered as lagged observations and errors.
“ar(1)ar(2)ar(3)ar(4)ar(5)ar(7)ar(8)ar(9)ar(12)ar(14)ma(1)ma(2)ma(3)ma(5)ma(7)ma(8)ma(14)ma(15)”.

In the estimated model for ARIMA model, from the above-mentioned factors, those factors would be valid that their p_value be less than 0.05 in 95% confidence level. After try and error process, the following valid components by using ARCH (1,1) method are found:
“ar(1)ar(2)ar(7)ar(8)ar(9)ar(14)ma(1)ma(3)ma(7)”.

The acquired model is nonlinear due to using ARCH method. These specified factors are recognized as lagged observations which are used as input parameters of the ANN model to create hybrid ARIMA-ANN structure.

4.3. Verification and Validation

In this section, the proposed hybrid ARIMA-ANN model is compared respect to ARIMA model and well-known regression’s models. The specified lag variables by ARIMA model are used as X-values in regression’s models. Indeed, the inputs of ANN are considered as X-values in regression’s models. The used regression’s models are described as follows:

Model (I) $$y = \alpha_0 + \sum_{i=1}^{7} \alpha_i X_i \quad (7)$$

Model (II) $$\ln y = \alpha_0 + \sum_{i=1}^{7} \alpha_i (\ln X_i) \quad (8)$$

Model (III) $$y = \alpha_0 + \sum_{i=1}^{7} \alpha_i X_i + \sum_{i=1}^{7} \beta_i X_i^2 \quad (9)$$

Model (IV) $$\ln y = \alpha_0 + \sum_{i=1}^{7} \alpha_i (\ln X_i) + \sum_{i=1}^{7} \beta_i (\ln X_i^2) \quad (10)$$

Model (V) $$y = \alpha_0 + \sum_{i=1}^{7} \alpha_i X_i + \sum_{i=1}^{7} \beta_i X_i^2 + \sum_{i}^{7} \sum_{i \neq j}^{7} \gamma_{ij} X_i X_j \quad (11)$$

Model (VI) $$\ln y = \alpha_0 + \sum_{i=1}^{7} \alpha_i (\ln X_i) + \sum_{i=1}^{7} \beta_i (\ln X_i^2) + \sum_{i}^{7} \sum_{i \neq j}^{7} \gamma_{ij} (\ln X_i X_j) \quad (12)$$

Eviews package and MATLAB software are used for implementing ARIMA process and regression's models, respectively. Table 2 shows the best results acquired by applied methods.

Table 2. The results of applied methods

Regression's models	MAPE (%)
Hybrid ARIMA-ANN	5.73*
ARIMA	14.0
Regression Model (I)	7.31
Regression Model (II)	7.14
Regression Model (III)	7.13
Regression Model (IV)	7.19
Regression Model (V)	6.63
Regression Model (VI)	7.19

As shown in Table 2, the ARIMA model is unable to forecast the considered time series for gasoline consumption. This observation could be resulted due to nonlinear behaviour of the considered time series. Although the regression's models dominate ARIMA model, their capability to forecasting the considered time series is less than proposed hybrid ARIMA-ANN model. Also the regression's model (V) is superior to other developed regression's models. Figure (8) shows the forecasted values of ARIMA process and different regression's models against actual values.

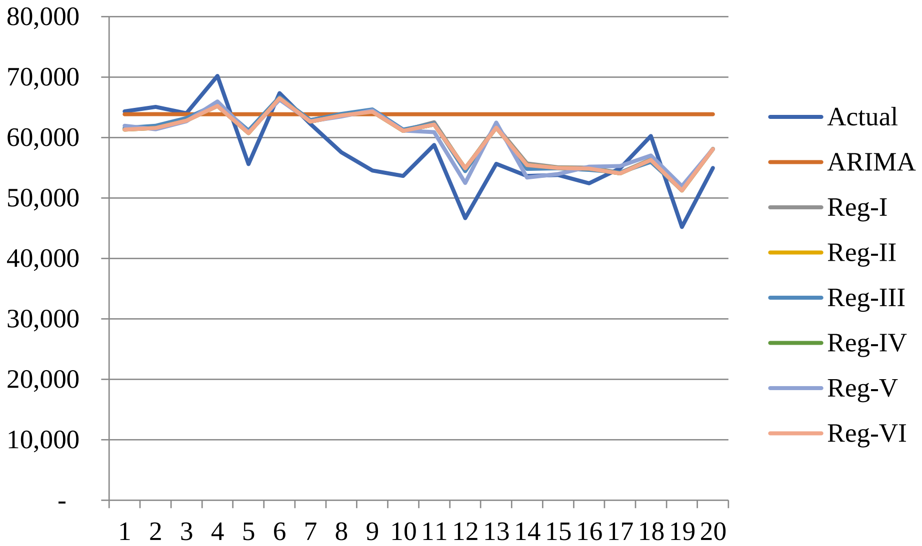

Figure 8. Forecasted values of different approaches

5. Conclusion

In this paper, we have proposed a hybrid ARIMA-ANN model for prediction of gasoline consumption. For this model a number of effective input data are extracted from the ARIMA process and applied in the structure of the ANN model. To show the applicability and superiority of the proposed framework actual data for robust period is used. MLP network is used and applied with the seven input variables. After tuning, the optimum number of neurons is determined in layers. The acquired results of the proposed hybrid ARIMA-ANN model are compared with different regression's models and ARIMA process. Although eliminating subside from gasoline price has led to appearing noisy data in gasoline consumption in Iran, the acquired results show high accuracy of about 94.27%. Also the results show that the ARIMA process is unable to model the nonlinear behaviour of time series even by using nonlinear ARCH method. Moreover, the proposed approach dominates well-known regression's models, which are commonly used in forecasting areas. It can be concluded from the acquired results that the proposed hybrid ARIMA-ANN model can be effectively used in gasoline consumption forecasting. Although the proposed approach is applied on a real case of gasoline consumption, it can be successfully used for other types of time series that have the similar structure.

The proposed model was implemented for predicting gasoline consumption In the National Iranian Oil Refining and Distribution Company (NIORDC). The results show the applicability and superiority of the proposed hybrid approach. Daily available data were collected for 7 years (2015–2021) in NIORDC. The prediction of gasoline consumption with high accuracy is crucial for Iran's National Oil Refining and Distribution Company (NIORDC) for several reasons. Accurate predictions help NIORDC allocate resources effectively, ensuring that refineries produce the right amount of gasoline to meet demand without overproduction or shortages. Understanding consumption trends allows NIORDC to make informed financial decisions regarding investments in refining infrastructure and distribution networks. Predictive analytics can streamline operations, reducing waste and improving the efficiency of the refining processes. This can lead to lower production costs and higher profitability. Accurate consumption forecasts can guide government policies related to fuel pricing, subsidies, and energy conservation measures, thus impacting overall economic stability. By predicting gasoline consumption, NIORDC can implement strategies to reduce emissions and promote cleaner energy alternatives, aligning with global environmental standards. Forecasting helps stabilize the market by anticipating fluctuations in demand, allowing NIORDC to adjust production and distribution strategies accordingly. Analyzing consumption patterns can provide insights into consumer behavior, enabling NIORDC to tailor its services and marketing strategies more effectively.

For future research, it is recommended to use other prediction methods and compare their results. It is recommended to use fuzzy logic to incorporate information that is presented in linguistic form.

References

Al Mamun, A., Sohel, M., Mohammad, N., Sunny, M. S. H., Dipta, D. R., & Hossain, E. (2020). A comprehensive review of the load forecasting techniques using single and hybrid predictive models. *IEEE access,* 8, 134911-134939. 10.1109/ACCESS.2020.3010702.

Austina, N., Senthilkumar, P., & Kanthavelkumaran, N. (2013). Artificial neural network involved in the action of optimum mixed refrigerant (domestic refrigerator).

Azadeh, A., Asadzadeh, S., & Ghanbari, A. (2010). An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: Uncertain and complex environments. *Energy Policy,* 38(3), 1529-1536. https://doi.org/10.1016/j.enpol.2009.11.036.

Azadeh, A., Babazadeh, R., & Asadzadeh, S. (2013). Optimum estimation and forecasting of renewable energy consumption by artificial neural networks. *Renewable and Sustainable Energy Reviews,* 27, 605-612. https://doi.org/10.1016/j.rser.2013.07.007.

Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G., & Moncada-Herrera, J. A. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment,* 42(35), 8331-8340. https://doi.org/10.1016/j.atmosenv.2008.07.020.

Fan, D., Sun, H., Yao, J., Zhang, K., Yan, X., & Sun, Z. (2021). Well production forecasting based on ARIMA-LSTM model considering manual operations. *Energy*, 220, 119708. https://doi.org/10.1016/j.energy.2020.119708.

Fathi, S., Eftekhari Yazdi, M., & Adamian, A. (2020). Estimation of contact heat transfer between curvilinear contacts using inverse method and group method of data handling (GMDH)-type neural networks. *Heat and Mass Transfer*, 56, 1961-1970. https://doi.org/10.1007/s00231-020-02832-x.

Forozandeh, M. (2021). The effect of supply chain management challenges on research and development projects using Fuzzy DEMATEL and TOPSIS approach. *Annals of Management and Organization Research*, 2(3), 175-190. https://doi.org/10.35912/amor.v2i3.801.

Ghahnavieh, A. E. (2019). Time series forecasting of styrene price using a hybrid ARIMA and neural network model. *Independent Journal of Management & Production*, 10(3), 915-933. https://doi.org/10.14807/ijmp.v10i3.877.

Ho, L. V., Nguyen, D. H., Mousavi, M., De Roeck, G., Bui-Tien, T., Gandomi, A. H., & Wahab, M. A. (2021). A hybrid computational intelligence approach for structural damage detection using marine predator algorithm and feedforward neural networks. *Computers & Structures*, 252, 106568. https://doi.org/10.1016/j.compstruc.2021.106568.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied soft computing*, 11(2), 2664-2675. https://doi.org/10.1016/j.asoc.2010.10.015.

Mbamalu, E. I., Chike, N. K., Oguanobi, C. A., & Egbunike, C. F. (2023). Sustainable supply chain management and organisational performance: Perception of academics and practitioners. *Annals of Management and Organization Research*, 5(1), 13-30. https://doi.org/10.35912/amor.v5i1.1758.

Melina, M., Sukono, S., Napitupulu, H., Mohamed, N., Chrisnanto, Y. H., Hadiana, A. I., . . . Nabilla, U. (2024). COMPARATIVE ANALYSIS OF TIME SERIES FORECASTING MODELS USING ARIMA AND NEURAL NETWORK AUTOREGRESSION METHODS. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 18(4), 2563-2576. https://doi.org/10.30598/barekengvol18iss4pp2563-2576.

Ong, C.-S., Huang, J.-J., & Tzeng, G.-H. (2005). Model identification of ARIMA family using genetic algorithms. *Applied Mathematics and Computation*, 164(3), 885-912. https://doi.org/10.1016/j.amc.2004.06.044.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing, volume 1: Explorations in the microstructure of cognition: Foundations*: The MIT press.

Tarafdar, H. M., & Ghadimi, N. (2014). Radial basis neural network based islanding detection in distributed generation.

Viana, C., Oliveira, S., & Rocha, J. (2024). Introductory Chapter: Time Series Analysis. *Time Series Analysis: Recent Advances, New Perspectives and Applications*, 3-13.

Wang, L., Zou, H., Su, J., Li, L., & Chaudhry, S. (2013). An ARIMA-ANN hybrid model for time series forecasting. *Systems Research and Behavioral Science*, 30(3), 244-259. https://doi.org/10.1002/sres.2179.

Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. *PhD thesis, Committee on Applied Mathematics, Harvard University, Cambridge, MA*.

Zahedi, M., Abbasi, M., & Khanachah, S. N. (2020). Providing a lean and agile supply chain model in project-based organizations. *Annals of Management and Organization Research*, 1(3), 213-233. https://doi.org/10.35912/amor.v1i3.440.