

A Survey of Different Prediction Models & Role of Artificial Neural Networks for Natural Gas Consumption

Prabodh K Pradhan¹, Sunil Dhal²

¹Research Scholar, Sri Sri University, Odisha, India

²Associate Professor, Sri Sri University, Odisha, India

Abstract: *It is very important fact that natural gas is now a very significant element of human beings day to day life. Local Distribution Companies (LDCs) are forecasting their customers' natural gas demand with different traditional and advanced methodologies. However the forecasting needs to be accurate to mitigate the demand and supply. A significant error on a single day can cost a lot to the customers of the LDC. Different attempts have been made to predict the natural gas prediction accurately by different research paper. A better and more realistic energy forecast is essential for the policy makers and LDC owners/managers when futuristic decisions for the next epoch. As this upsurge in energy usage may posture a threat to economic development in the nation, Therefore, the policy makers need an accurate prediction of natural gas consumption to meet the demand and needs of the people. This paper looks at the brief description of each method and techniques described in different research papers.*

Keywords: Neural Network, prediction, ARIMAX, Gas consumption, Mathematical model

1. Introduction

It is a fact that, natural gas is now an essential part of social development and is a key factor to improved life[1]. Also energy is considered to be a significant input in economic development and a prime factor in the generation of wealth [4]. In view of the above considerations, Taking the above indicators as, many nations have considered natural gas and energy as vital necessities and had designed their energy policies to provide the natural gas energy on a timely, reliable, cost-effective, environmentally friendly, and high-quality basis so as to aligned to the development and social progress. In[4] author states that the term “energy security” is rising higher on the agenda of policymakers and city administrators in Turkey, European Union and the Balkan countries.

Worldwide, most of the increase in the demand for natural gas is expected to come from developing countries, where gas consumption is expected to grow from 0.5 tcm in 1999 to 4.5 tcm in 2020. Following the economic crises worldwide, on the assumption that the world economy began to recover in 2010, primary gas demand is projected to rebound, growing on average by 2.5% per year between 2010 and 2015. Gas demand in the OECD countries is estimated to increase through 2030, as coal- and fuel-oil-fired generators are replaced by plants using renewable and nuclear power [5]. The use of natural gas in OECD countries is projected to grow by 2.4% annually, compared to a more modest rate of increase of 1.1% in oil consumption. This would account for 49% of the projected increase in total energy consumption of these countries [4].

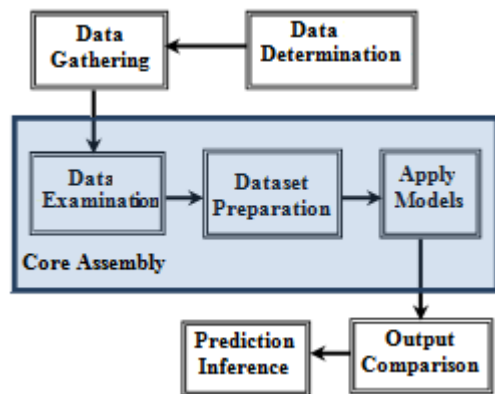
Literature review shows that the artificial neural networks (ANN) models have been used in different fields such as psychology, medicine, mathematics, engineering, finance, and economics[6]. As the literature shows, the applications

of the ANN models in economics include sectors such as water resources, tourism and energy sector [6].

In [7], It has been discussed that the prediction of natural gas consumption is crucial for the gas distribution and transportation companies as well as for the government agencies associated to the natural gas sector. In particular, the short range prediction, 1 to 5 days, is important to ensure the normal supply of natural gas to a given city or community. This type of prediction is particularly important for countries like Argentina, where the production sites are far from the major centers of consumption. In the case of Argentina these distances are of about 2000 km. There is also a need to predict the firm component of consumption in the intermediate range of time, within 1 to 4 years, in order to adapt and upgrade the infrastructure of transportation and distribution. This type of prediction is also useful for all the sectors of the gas industry that need to plan their production and optimize their anticipated purchase. The most important factors that affect the gas consumption of residential and commercial users are temperature, day-of-the-week ~holiday or working day! and prevailing scenario of consumption. Other factors that may also influence the consumption are: wind speed and its direction, humidity, etc. Due to the lack of reliable information on these parameters, we have not included them in our model.

2. Generalized Prediction Flow

The Prediction of Natural Gas flow passes through different stages as shown in below figure. The short description of each stage is described below.



Data Determination: IN general, for any prediction model, right data properties to be chosen before proceeding for the other stages. Some of the few examples are, temperature, cultural event etc. are some of the important factors which to be taken into consideration. If the other properties are included in the prediction model, then it will deviate from the actual results, which will lead to an erroneous prediction.

Data Gathering: Data gathering is another important stage where the accurate data set has to be gathered for the reliable source. For example, the data set collected from the non-reliable source like temperature captured from a faulty device from public place may contain erroneous data and would not give the correct result.

Data Examination: In Data Examination stage, the gathered data is examined to reduce some of the error prone data. The redundant data and invalid data. For example, the temperature would not be less than 0 degree centigrade or more than 50 degree centigrade, which is impractical. This will reduce the prediction complexity.

Dataset preparation: Data set to be accurately prepared for learning and training. The other set of data to be prepared and kept reserved for evaluating the model. The more close output resulted after the prediction mode is applied, the better model it assumed to be.

Apply Model: After all the above stage completed, the prediction model is applied and the output data is now ready for comparison.

Output Comparison: In this stage, the output data resulted from the Core Assembly system, would be compared with the dataset reserved for comparison and the %match or error calculated to verify the closeness or accuracy.

Prediction Inference: The more accurate result will infer the model reliability and trust worthiness. If the model provide higher rate of error, then again the model and data set is re-evaluated as per the above process and the system to be corrected.

2.1 Prediction Models

There are multiple prediction models have been evolved to evaluate the correct prediction of natural gas. These models are shown as in below figures.

2.1.1 Mathematical Models:

The most common mathematical modeling techniques used to forecast daily demand are multiple linear regression and artificial neural networks.

a. Multiple Linear Regression Model :

Multiple linear regression Multiple Linear Regression (MLR) [8,9] is one of the most commonly used methods for prediction models, and it has been applied to utility forecasting [10]. Some of the below models discussed in [3].

1. OLS regression Model

In a time series context, the OLS regression model takes the form of:

$$y_t = \beta X_t + \epsilon_t \quad (1)$$

Here, left-hand-side variable y is the dependent variable and right-hand-side vector X includes independent variables. Normally, vector X includes variables that are considered to be factors that influence dependent variable y . Within the context of our time series approach in this paper, however, X also includes lags of dependent variable y .

2. ARMAX Model

Autoregressive time series models can be structured as univariate or multivariate models. In a univariate specification called the autoregressive integrated moving average (ARIMA), future forecasts of a variable are based on the historical values of the same variable. ARIMA specification takes the following general form:

$$y_t = a_0 + a_1 y_{t-1} + \dots + a_p y_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} \quad (2)$$

Here, there are potentially p autoregressive (AR) and q moving average (MA) terms. p and q are integers and are called orders of the model. a_i ($i = 0, 1, 2, \dots, p$) and b_i ($i = 1, 2, \dots, q$) are parameters to be estimated. Disturbances ϵ_t are assumed to be independent and identically distributed (iid) and normally distributed with mean zero and variance σ^2 .

In many cases, in addition to the lags of the dependent variable, it is useful to include independent factors in the specification that are considered to influence the dependent variable. Since natural gas consumption is forecasted in this paper, it is natural to include other independent variables such as temperature, price of natural gas, and number of gas consumers. Fortunately, the statistics package that was used allows the specification of a structural model with AR and/or MA disturbances. This type of a specification is called an ARMAX model. Formally, an ARMAX model can be expressed as a structural equation and an equation to specify ARMA disturbances, respectively [11]:

$$y_t = X_t \beta + \mu_t \quad (3)$$

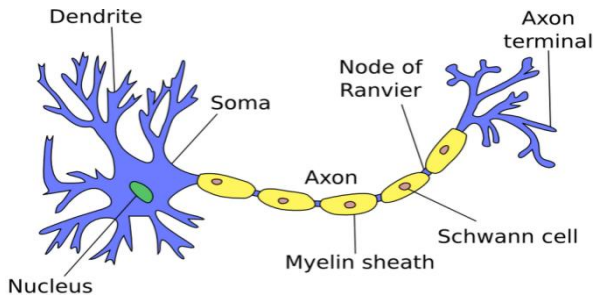
$$\mu_t = a_1 \mu_{t-1} + \dots + a_p \mu_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} \quad (4)$$

Disturbances of the structural equation μ_t are allowed to follow an ARMA (p, q) process. Likewise, ϵ_t values are assumed to be iid and normally distributed with mean zero and variance σ^2 . Vector X_t includes both independent factors and possibly lags of the dependent variable. This specification is flexible enough to exploit all available

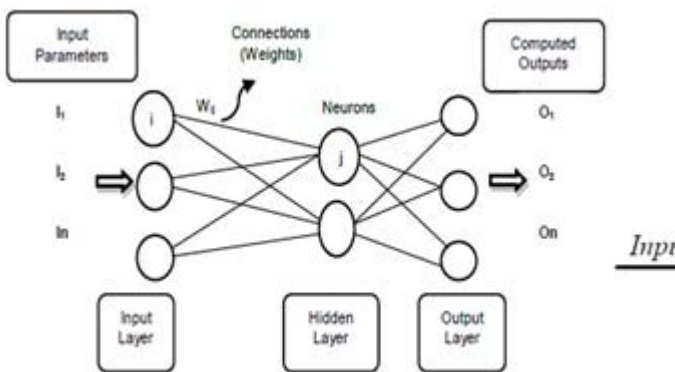
historical information content for forecasting future values of the dependent variable.

2.1.2 Artificial Neural Network Model

Artificial neural networks Artificial Neural Networks (ANN) [7] are mathematical models which can approximate any (nonlinear) continuous function arbitrarily well. ANNs are modeled by using neurons resembling those of the human brain. A neuron, fundamental processing element of an ANN, is shown in below figure.



Initially, the data to be taught are entered into a program based on ANN system. Then, the program models these data through different neuron and network structure. It aims to produce values close to the value, which is targeted to be taught, by using neuron weights. The ANN acquires knowledge through a training process. Modelers of gas consumption have been attracted to ANN's because of this capability of mapping unknown nonlinear relationships between inputs and the output. In particular, the nonlinear properties of the ANN allow the direct input of temperature, wind speed, and prior day temperatures into the ANN nodes without accounting for interactions and the nonlinear response of these impacts. In addition, the training process builds an input-output relationship that interpolates well to a situation that may not exactly match the training data. However, while an ANN is quite good at interpolating a solution that was not presented during training, it is not as good at extrapolating outside the domain of the training knowledge.



It was shown [14] that, at least theoretically, a single hidden layer is always used quite safely for the majority of practical problems. Such three-layer ANN is shown in Figure 2. In addition, more than two hidden layers should not be needed in practice. The network weights are adjusted by training the network (Figure 3.). It is said that the network learns through examples. The idea is to give the network input signals and desired outputs. To each input signal the network produces

an output signal, and the learning aims at minimizing the sum of squares of the differences between desired and actual outputs. The learning is carried out by repeatedly feeding the input/output patterns to the network. One complete presentation of the entire training set is called an epoch. The learning process is usually performed on an epoch-by-epoch basis until the weights stabilize and the sum of squared errors converges to some minimum value. The most often used learning algorithm for the ANN is the backpropagation algorithm. A more powerful algorithm is obtained by using an approximation of Newton's method called Levenberg-Marquardt. In applying the algorithm to the network training, the derivatives of each sum of squared error (i.e. with each training case) to each network weight are approximated and collected in a matrix. This matrix represents the Jacobian of the minimized function. Levenberg-Marquardt algorithm is particularly suitable in solution of approximation problem. In essence, the learning of the network is nothing but estimating the model.

For example, in the gas estimation problem, this means that if the ANN model was not trained with historical data from days of extreme weather, the model may not perform well on such days. GasDay™ uses a separate ANN for each forecasting horizon.

2.1.3 Dynamic Model

Combining multiple forecasts from models such as artificial neural networks or multiple linear regression can reduce errors arising from faulty assumptions, bias, or mistakes in data [7]. Bates and Granger suggest that combining several forecasts together tends to decrease forecasting error because the combined forecast has equal to or frequently less variance than each of the component forecasts, and Dickinson [13] provides a mathematical proof of this. Armstrong [12] surveyed research on combining forecasts over the last 40 years, concluding that to obtain the best combined forecast accuracy the following guidelines should be considered.

- use different component forecast methods,
- use at least five component forecasts when possible,
- use equal weights unless you have strong evidence to support unequal weighting of forecasts,
- use trimmed means,
- use different data, and
- use the track record and domain knowledge to vary the combination weights.

Natural gas forecasting is an ideal case for combining forecasts, because the forecaster is not always certain which forecasting model is most accurate. As discussed earlier, linear regression extrapolates better than ANNs, but the ANNs often perform better on days similar to ones in its training set. While our two component models (MLR and ANN) each produce estimates of consumption, a weighted combination of these models often yields improved results over the best of the component model estimates. The combination of component model estimates also helps to hedge the forecast since it tends on average to produce estimates that deviate less from the actual consumption.

2.2 Factors that affect daily demand

Temperature: The most significant factor for modeling natural gas consumption is temperature, since most gas is used for space heating.

Wind Effects: Another important factor is wind, because buildings lose more heat on a windy day than on a calm day. Wind could be added as another term to the models above, but then the wind effect would be the same at all temperatures, while it is well known that the impact of wind increases with HDD.

Previous day demand: Typically, the load forecasts are made for the coming day before the current day's gas day is complete. Thus, the current day's demand is not known. However, yesterday is over, so the forecast for that day may be known. Adding this and earlier daily loads as inputs to the forecast model, making it autoregressive, can reduce forecast error significantly.

Day of the week effects: Gas consumption varies by the day of the week. For example, on weekends, as residential consumption increases, demand is typically more than offset by decreased consumption of both commercial and industrial consumption. Gas load forecasters have used many techniques to try to capture this effect, primarily by adding day-of-the-week indicator model inputs.

Holidays and days around holidays: Holidays and days near holidays typically have lower demands than if the day was not a holiday. One approach that can be used with or without the above mentioned day of week adjustment is to average the residual errors in the training data on specific holidays and adjust the demand forecast. For example, if, after parameterizing our model, we evaluate the model on all of the New Years Days in the training set and calculate the forecast error as the demand forecast minus the actual flow, and calculate the average errors, we can subtract this average error to adjust the forecast on New Years Day.

Other Factors : Many other potential factors exist, such as solar radiation, cloud cover, precipitation, dew point, direction of the wind, tap water temperature, bill shock, occupancy rates, industrial production rates, and other econometric factors, to name a few. Some of these factors can be measured directly, while others cannot, or at least, cannot be measured reasonably. Solar radiation and cloud cover affect temperatures throughout the day. For example, the evening temperature decrease is less on a cloudy day than on a clear day. Wind direction has an effect, especially in coastal regions next to oceans or the Great Lakes. The dew point is a measure of humidity, and more gas tends to be consumed on humid days. Precipitation measures rain and snow fall.

3. Conclusion

Possibility of artificial neural network utilization for forecasting of natural gas consumption is shown in this paper. Specified consumption area is analyzed and appropriate training set, which includes historical weather and consumption data, is defined. Parameters of ANN are

obtained using a Levenberg-Marquardt training algorithm. Analyses of results obtained for training and test sets show that designed artificial neural network could be useful for natural gas consumption forecast problem.

References

- [1] Omer F, Selim Z, Ahmet C, ALIS, KAN3, Pinar O, "Forecasting natural gas consumption in Istanbul using neural networks and multivariate time series methods" Turk J Elec Eng& Comp Sci, Vol.20, No.5, 2012, pp 695 - 711
- [2] M.A. Kili, c, "Turkey's natural gas necessity, consumption and future perspectives", Energy Policy, Vol. 34, pp.1928-1934, 2006.
- [3] M. Tun, c, " U. C, amdali, C. Parmaksizoğlu, "Comparison of Turkey's electrical energy consumption and production with some European countries and optimization of future electrical power supply investments in Turkey", Energy Policy, Vol. 34, pp. 50-59, 2006.
- [4] Bassam M. AbuAl-Foul, Forecasting Energy Demand in Jordan Using Artificial Neural Networks, Topics in Middle Eastern and African Economies Vol. 14, September 2012, pp. 473 – 478
- [5] S. Gill, J. Deferrari "Generalized Model of Prediction of Natural Gas Consumption" Transactions of the ASME Vol. 126, JUNE 2004 pp. 90 – 98
- [6] N. R. Draper and H. Smith, Applied Regression Analysis, 3rd ed., John Wiley & Sons, New York, 1998.
- [7] A. S. Goldberger, Topics in Regression Analysis, Macmillan, New York, 1968.
- [8] T. Haida and S. Muto, Regression based peak load forecasting using a transformation technique, IEEE Trans. Power Syst. 9(4) (1994), 1788{1794.
- [9] StataCorp, Stata Statistical Software: Release 11, College Station, Texas, StataCorp LP, 2009.
- [10] J. S. Armstrong, Principles of Forecasting a Handbook for Researchers and Practitioners, Kluwer Academic Publishers, Boston, 2001.
- [11] J. P. Dickinson, Some statistical results in the combination of forecasts, Oper. Res. Q. 24(2) (1998).
- [12] Yalcinoz, T. and Eminoglu, U.: Short term and medium term power distribution load forecasting by neural networks, Energy Conversion and Management, Vol. 46, No. 9-10, pp. 1393-1405, 2005.