yönetim ve ekonomi araştırmaları dergisi
journal of management and economics research
Cilt/Volume: 21 Sayı/Issue: 3 Eylül/September 2023 ss. /pp. 312-332
E. Avcı http://dx.doi.org/10.11611/yead.1323635

IMPROVING TURKEY'S NATURAL GAS DEMAND FORECASTS: A DATA ANALYTICS APPROACH

Asst. Prof. Ezgi AVCI (Ph.D.)*

ABSTRACT

As the world population and consequently energy demand is rapidly increasing, primary energy sources are needed for sustainable development. One fifth of the world's energy needs are provided by natural gas, due to it being the cleanest burning fossil fuel. However, as a fossil fuel, natural gas is a limited resource and requires efficient use. Therefore, planning for natural gas consumption is of great importance. The most critical input for this planning is natural gas demand forecasts. Due to the increasing importance of the subject, numerous studies on natural gas demand forecasting have been conducted in the literature. However, the scientific studies carried out in Turkey are quite limited. In this article, the current demand forecast situation in Turkey is revealed, and a data analytics framework based on big data is developed and proposed.

Keywords: *Natural Gas Demand Forecast, Big Data Analytics, Decision Support Systems, Policy Design.*

JEL Codes: *C10, C55, C80.*

TÜRKİYE'NİN DOĞALGAZ TALEP TAHMİNLERİNİN İYİLEŞTİRİLMESİ: VERİ ANALİZİ YAKLAŞIMI

ÖZET

Dünya nüfusu ve dolayısı ile enerji talebi hızla artarken, sürdürülebilir kalkınma için birincil enerji kaynaklarına gerek duyulmaktadır. Dünya enerji ihtiyacının beşte biri, en temiz yakımlı fosil yakıt olmasından dolayı, doğalgazdan sağlanmaktadır. Doğalgaz fosil yakıt olması dolayısı ile sınırlı bir kaynaktır ve verimli kullanım gerektirmektedir. Bu nedenle doğalgaz tüketim planlaması büyük önem arz etmektedir. Söz konusu planlamanın en önemli girdisi doğalgaz talep tahminleridir. Konunun artan öneminden dolayı, literatürde çok sayıda doğalgaz talep tahmin çalışması yapılmıştır. Türkiye de yapılan bilimsel çalışmalar oldukça sınırlıdır. Bu makalede doğalgaz talep tahmini ile ilgili dünya

* TED University, Graduate Programs Instıtue, Applied Data Science Master Program Master of Science Program in Economics Faculty, Ankara/ Türkiye, E-mail: ezgi.avci@tedu.edu.tr

Makale Geçmişi/Article History

Başvuru Tarihi / Date of Application	: 6 Temmuz / July 2023
Düzeltme Tarihi / Revision Date	: 20 Eylül / September 2023
Kabul Tarihi / Acceptance Date	: 25 Eylül / September 2023

Araştırma Makalesi/Research Article

literatürü detaylı bir şekilde irdelenecek, Türkiye'deki mevcut talep tahmin durumu ortaya konacak ve büyük veriye dayalı veri analitiği yaklaşımı önerilecektir.

Anahtar Kelimeler: *Doğalgaz Talep Tahmini, Büyük Veri Analitiği, Karar Destek Sistemi, Politika Tasarımı.*

1. INTRODUCTION

Energy plays a vital role in the social, economic, and environmental aspects of sustainable development for nations. Energy is directly/indirectly related to industrial production, agricultural output, health, access to water, population, education, quality of life, and transportation (e.g. Xiong et al., 2014). While global energy demand is rapidly increasing, primary energy sources are needed for sustainable development (e.g. Fan and Xia, 2012; Panella et al., 2012; Gracias et al., 2012; Aydin, 2015; Azadeh et al., 2015; Szoplik, 2015). One fifth of the world's energy demand is met by natural gas, due to its clean combustion as a fossil fuel. As a fossil fuel, natural gas is a limited resource and requires efficient use. Thus natural gas consumption planning is of great importance. The most critical input of this planning is natural gas demand forecasts. In this article, the current demand forecast situation in Turkey is revealed, and a data analytics approach based framework on big data is developed and proposed.

In Turkey, natural gas demand forecasts are used in;

- inputting into Turkey's energy policies, objectives, strategies, and plans,
- determining Turkey's annual investment programs,
- defining natural gas purchase and sales strategies,
- carrying out gas pull programs based on main output points,
- conducting monthly and annual capacity reservation, idle capacity, capacity transfer, and capacity transfer transactions,
- performing capacity reservation transactions and daily program notifications,
- determining the offer price of balancing gas on a monthly basis,
- in gas master plan studies,
- planning the investment of natural gas transmission lines and efficient operation,
- planning for spot LNG needs arising from winter conditions,

Due to the increasing importance of the subject, many natural gas demand forecast studies have been carried out in the literature (e.g. Soldo, 2012; Voudouris et al., 2014; Fagiani et al., 2015; Khan, 2015; Suganthi and Samuel, 2012). Natural gas demand forecasting has been researched at many different levels are summarized in Table 1 (Soldo, 2012; Tamba et al; 2018).

Table 1. Levels of Natural Gas Demand Forecasting

Forecasting area	Forecast approach	Forecast horizon	Used daily demand
• Global level • National level • Regional level • Gas distribution system-level • Individual user consumer-level	• Bottom-up • Top-down	• Hourly • Daily • Monthly • Annually • Combined	• Hourly • Daily • Monthly • Annually

2. LITERATURE

In this section, the literature on natural gas demand forecasting is evaluated primarily according to the forecasting area (USA, Europe, and Turkey), and then according to the forecasting method used. Although there are many academic studies for the USA and Europe, examples have been given due to the page limit. The existing literature for Turkey has been taken.

2.1. Studies in America

Studies using econometric methods include: Balestra and Nerlove (1966) made commercial and residential demand forecasts using econometric parameters and the Least Squares (LS) method. They observed their models on six-year data from 36 states. The results obtained have shown that time-invariant regional effects constitute approximately three-quarters of the total residual variance of the demand equation. Berndt and Watkins (1977) developed an econometric model for commercial and residential demand forecasting in British Columbia and Ontario. With this model, they explained most of the demand fluctuations observed in these states in the past as a function of price, household income, and temperature. It was observed that heating degree days have an effect of approximately 0.75 on the elasticity of natural gas demand. On the other hand, while prices have a significant effect on long-term natural gas demand, they do not have a significant effect on short-term natural gas demand. A 10% price decrease increased natural gas demand by 5% in both states over a seven-year period. Maddala et al. (1997) applied the classic applied Bayes procedure to the problem of estimating short and long-term household electricity and natural gas demand elasticity in 49 states across America. Here they encountered three alternatives. The first is to use separate time series regressions for each state. This option has not been seen as a good option as it is difficult to interpret and gives wrong signals. The second is to group the data and use panel data estimators. This approach also did not find a good response, as the hypothesis of homogeneity of coefficients was rejected. The third option is to use shrinkage estimators. This option gave more reasonable results. Baltagi et al. (2002) reached the best out-of-sample performance for American states by using homogenous panel data estimators. It was seen that when these data were used in the prediction of inter-state heterogeneous models, the worst out-of-sample performance was presented by individual predictions. Shrinkage estimators, although providing better performance than individual predictions, have failed against simple homogeneous panel data predictions in out-of-sample predictions.

Payne et al. (2011) modeled the household natural gas demand in the state of Illinois using the Autoregressive Distributed Lag (ARDL) method. In this study, short and long term demand dynamics were simultaneously examined, research was conducted on short and long term income, price and cross price elasticity, the effect of the heating degree day variable on household natural gas demand was determined, and the performance of the model used in tracking real household natural gas demand data was observed. An annual data set for the period 1970-2007 was used. With the ARDL method, a balance relationship was observed in the long term between per capita household natural gas demand values, real household natural gas prices, real per capita income, real household electricity prices, real fuel prices, and heating degree days. The coefficients of the long-term balance relationship showed that only real household natural gas prices, real household electricity prices, and heating degree days are statistically effective at least at a rate of 10%. In addition, the ARDL model performed well in following real household natural gas demand values in the state of Illinois. Kalashnikov et al. (2009) performed regression curve completion analysis with monthly price and demand data for American states during the period January 1989 – December 2007 and created a classification of the states in terms of household demand and price. One of the most significant problems for the American market is defining which regions have similar economic characteristics. One of the claims of the authors has emerged in this direction: If it can be determined which regions have similar price and demand functions and a classification can be made in this direction, the number of variables needed in scenario analysis can be significantly reduced. In many past studies related to natural gas demand and price, linear logarithmic transformations have been applied, the main reason for this is that elasticity calculations are a very important part of econometric analysis and logarithmic functions facilitate the finding of numerical elasticity values.

Studies conducted using artificial intelligence methods: Lim and Brown (2001) used genetic algorithms to find the inputs of the natural gas hourly demand forecast model. Then, they put forward 23% less average error than the current method with linear regression-based models. The data set consists of real distributed natural gas data and real weather data (wind speed and temperature) from June 1998-August 2000 period. Pang (2012) improved demand forecasts using cloud cover, precipitation, and dew point data as independent variables. Precipitation generally provided a 2.5% improvement in the model, cloud cover improved the root mean square error (RMSE) by 1.2% as an indicator of radiation effect, and dew point provided a 2.4% RMSE improvement as a humidity indicator. These 3 parameters provided an average improvement of 5% RMSE compared to the existing model. An improvement study was also carried out using wind direction data, but it was seen that adding inputs according to the directions did not have a significant effect, therefore it was not included in the model. Merkel et al. (2018) used a deep neural network model for short-term natural gas demand forecasting. The obtained results were compared with 62 data sets obtained from many different regions of America. Each data set consists of 10 experimental sets and 1-year test set. This new approach is 9.83% more successful in terms of weighted mean absolute percentage error (WMAPE) than neural networks run with old methods. The authors claim that deep neural network algorithms are more successful in short-term forecasts than linear regression methods and artificial neural network algorithms.

2.2. Studies in Europe

Studies using econometric methods: Gutierrez et al. (2005) used the Gompertz-type stochastic growth model to estimate Spain's natural gas demand (annually) from 1998-2000. The results of the study were compared with two different stochastic growth models, and the Gompertz-type stochastic growth model produced better prediction results than the other two models. Pappas et al. (2008) used the ARMA model to predict Greece's daily electricity demand between 2004-2005. After the data set was purified from seasonality, predictions were made by choosing the ARMA model with the highest fit level among the multiple ARMA models created.

Studies conducted using artificial intelligence methods: Ivezic (2006) used artificial neural networks to predict the short-term natural gas demand of Belgrade. In this study, the predicted maximum and minimum temperatures for the day to be predicted, the natural gas consumption data of the past days, and dummy variables were used as model inputs. The dataset covers the dates between 01/10/2001 - 31/12/2004. This study suggested that artificial neural networks are a promising model for natural gas demand predictions. Potocnik et al. (2007) predicted Slovenia's daily and weekly natural gas demand using parametric and non-parametric methods with past natural gas usage, past weather data, weather forecast data, and dummy variables. The data set used in the study covers the period from June 2002 to December 2004. Beccali et al. (2004) estimated Palermo's (Italy) short-term electricity demand between 2001-2003 through an artificial neural network model using temperature, relative humidity, and solar irradiation as inputs. The developed hybrid model consists of a combination of supervised and unsupervised artificial intelligence algorithms. The results have shown that this model is a valid model that can be used in short-term demand forecasting problems. Vitullo et al. (2009) used temperature and past temperature values, wind speed, heating degree day value and its lagged values, wind-corrected heating degree day value and its lagged values, past demand value, and dummy variables as inputs for the linear regression and artificial neural networks models they used for natural gas demand forecasting in their study.

2.3. Studies in Turkey

They are quite limited and are summarized below:

Studies conducted at the beginning of the 2000s used the degree-day method. Durmayaz et al. (2000) used the degree-day method to predict seasonal natural gas demand for Istanbul in their study. In this study, a fuel consumption model with different building materials and the number of people living was established as a case study of a prototype building. It was assumed that this prototype building was located next to the Göztepe meteorological station on the Asian side of Istanbul. The outside air temperature data recorded at the Göztepe meteorological station between 1990-1997 were used to calculate the hourly averages of the outside air temperatures for each day during the heating season in those years. It was observed that the heating season started on the 292nd day and lasted for 201 days until the 127th day of the year. As a result, seasonal natural gas demands were revealed according to different building conditions and different numbers of people living in the apartment. Gumrah et al. (2001) used the degree-day method to model natural gas demand for Ankara. The data include gas demand measured between 1991-1997, the annual number of subscribers, degree-day values, changes in the number of subscribers, and usage rates per subscriber. A demand model was established as a function of the number of subscribers and weather conditions. It was concluded that Ankara's demand is highly dependent on weather conditions and the number of subscribers, and they confirmed with Fourier series that demand models using the degree-day method are reliable. Sarak and Satman (2003) used the degree-day method to estimate the demand for natural gas used for domestic heating in Turkey. Cities near existing and planned natural gas pipelines were selected for this study. The independent variable dataset includes heating degrees, population, and settlement records. The fuel consumption model they used to determine domestic natural gas demand nationwide includes: heating degree-day, heat transfer coefficient for the prototype building, fuel heating value, and efficiency of the heating system. The results were used to identify the energy demand distribution and select areas where consumption is highest.

Aras and Aras (2004) applied first-degree autoregressive time series models to predict domestic natural gas consumption. The data include monthly domestic natural gas consumption and daily average temperatures for a 61-month period between 1996-2001 for Eskişehir. Due to the basic usage of natural gas changing throughout the year, they divided the data into two periods, heating and non-heating. They created three first-degree autoregressive models where the deterministic component is a periodic function of time and degree-day values. They found that using separate models for each period, instead of using a single model throughout the year, resulted in a significant reduction in prediction errors. Görücü (2004) conducted a multivariate regression analysis for Ankara's natural gas consumption prediction. Independent variables taken were degree-day for heating, TL/USD exchange rate, number of subscribers, and gas sales price. Two models were developed; the first model developed with data between 1994 and 1997, and the second with data from 2001. The main distinction between the two models is that the second model was developed using data during crisis times and is used for predictions when there is a crisis. Özmen et al. (2018) used MARS and CMARS models for Ankara's domestic natural gas consumption prediction. In addition to meteorology and consumption data, gas unit cost, TL/USD exchange rate data were used. The data sets cover the years 2009-2013. The period 2009-2012 was selected as the training set, while the year 2013 was used as the test set. The variables used in the study are: heating degree-days, natural gas consumption, natural gas price, and USD/TRY exchange rate. The MARS and CMARS models were developed and compared. It was found that the MARS model had better forecasting performance.

Studies conducted using artificial intelligence methods are summarized in this chapter.

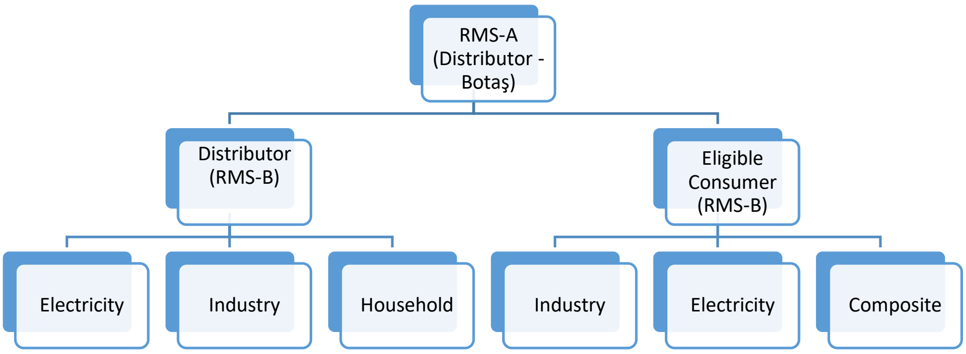
Kızılaslan and Karlık (2009) used the daily natural gas consumption in Istanbul for prediction. They utilized data from 2004 to 2007, including daily natural gas consumption, maximum and minimum temperatures, the previous day's maximum and minimum temperatures, the number of consumers, and the previous day's consumption amount. All these variables were found to be statistically significant. Dombaycı (2010) used an ANN (Artificial Neural Network) to predict the hourly energy consumption of a house representing Denizli province. The model was trained with heating energy consumption values from 2004-2007 and tested with 2008's heating energy consumption values. Taşpınar et al. (2013) used SARIMAX, ANN-MLP, ANN-RBF, and multivariate OLS for Sakarya's natural gas consumption forecast. The dataset included Sakarya province measurements from 2007-2011. The independent variables used were ambient temperature, average cloud cover, relative humidity, wind speed, and atmospheric pressure. The SARIMAX model was found to outperform the other models, and ambient temperature and cloud cover were identified as the most impactful variables on consumption. Demirel et al. (2012) attempted to predict Istanbul's daily natural gas consumption using daily consumption, average temperature, and DG price data from 2004-2009 via OLS, ARMAX, and ANN methods. They found that the ANN performed better than the other methods. Kaynar et al. (2010) made weekly predictions using Turkey's weekly DG consumption for the years 2002-2006. No explanatory variable was used. All models were found to perform similarly.

As seen from the literature review, academic studies conducted in Turkey have been carried out based on specific provinces and with limited data. As a result, these studies tend to be methodologically weak due to lack of feature engineering, parameter finetuning and lack of variety of compared models and do not offer opportunities for comparison and verification.

3. CONCEPTUAL BACKGROUND

BOTAŞ transmits natural gas to many different locations in Turkey, and these transmission points are classified as main and secondary. There are a total of 437 main exit points and 724 secondary exit points. The demand data read from these points provide us with information about different consumers' natural gas usage. Consumers who draw natural gas from the main exit points are classified as eligible consumers and distribution companies. It should be noted that while eligible consumers might operationally draw natural gas through a distribution company, they may not be commercially connected. The relationship between the exit points is explained in Figure 1.

Figure 1. Relationship Between Turkish Natural Gas Transmission Outlets



Eligible consumers; depending on their natural gas usage purposes; are divided into industry, electricity, and composite consumers. While the industrial consumer uses natural gas for manufacturing purposes; the electric consumer uses natural gas for electricity generation. The composite consumer uses natural gas for both manufacturing and electricity generation.

Distribution companies; distribute the natural gas from the exit points to their own consumers. The consumers of the distribution company are also divided into industry, electricity, and residential, according to their natural gas usage purposes.

The appropriate collection, classification, and storage of all consumers' natural gas consumption data is of great importance for the analyses to be performed on these data.

BOTAŞ's current data management involves classifying and storing the data collected daily from the exit points through the Microsoft Office Excel program. Primarily, the collected consumption values report is automatically classified for all consumer types via the Excel program.

The consumption data at the main exit points, named as RMS-A (Regulating and measuring station), and the consumption data at the secondary exit points, named as RMS-B, are first gathered in a daily consumption report called "EBT Point Allocation Quantity List". This report is prepared daily and includes the consumption data of all points in Turkey, information about which region and city this point is in, information about which distribution company operates it, the code of the point, and information about which consumer uses the allocated amount at the point and for what purpose. Then, these pieces of information are sent to a different Excel table for the different consumer types mentioned above and classified. After this classification process, detailed consumption information, such as the total consumption at the RMS-A points and, if any, how much of this consumption was used by different types of consumers with the RMS-B points, is segregated.

There may occasionally be errors in this data. These errors can arise either due to incorrect consumption readings by the distributor companies (Type 1 error), or due to technical errors at the exit points (Type 2 error). In the current system, after the data is classified, a graphical control method is applied for Type 1 errors. If there is a sharp movement outside of normal in the comparative consumption graph, it is checked whether this is caused by an error. Then, the necessary changes are made and corrected in the detected erroneous data. The change method here; is to find the erroneous consumption value with classic arithmetic methods or to use it as a substitute with the assumption that the previous week's consumption (especially industrial consumption) will have an equivalent consumption value under normal conditions.

4. PROPOSED DATA ANALYTICS APPROACH

We propose a data analytics framework that adopt both the bottom-up forecasting approach based on the engineering perspective and the top-down approach based on the economic perspective for natural gas demand forecasting (van Beeck, 1999). Initially, consumer segments is formed as Electricity, Residential and Industry, then Electricity and Residential Sector Natural Gas Consumption Forecast Models are established. On the other hand, using the Natural Gas Consumption data of the Electricity and Residential Sector as external data, the Total Natural Gas Consumption Model of Turkey is created. Finally, by subtracting the electricity and residential sector consumption forecasts from the total natural gas consumption forecast of Turkey, the Industry Sector Consumption Forecast is formed.

4.1. Electricity Sector Natural Gas Consumption Forecast:

When forecasting the total daily natural gas consumption of the natural gas power plants in the electricity sector, firstly, the merit order of electricity supply sources is arranged. In this sequence, the supply source with the highest unit cost, which is located at the end and whose production will be affected by others, is modeled. As the natural gas power plants currently have the highest unit cost in the merit order, Turkey's total electricity consumption is primarily modeled. Then, electricity production models of other supply sources excluding natural gas are established. Finally, by subtracting the total electricity production forecasts of other supply sources from the total daily electricity forecast of Turkey, the amount of electricity to be produced by Natural Gas Power Plants is determined. This amount is converted into Natural Gas volume using the average efficiencies of Natural Gas Power Plants and the conversion formulas of kWh to Sm3.

The classification of electricity production supply sources excluding Natural Gas and their plant counterparts is as follows:

- Run-of-river Hydroelectric Power Plant (Stream HPP)
- Dam Hydroelectric Power Plant (Dam HPP)

- Asphaltite Coal Power Plant
- Biomass Power Plant
- Fuel Oil Power Plant
- Solar Power Plant
- Imported Coal Power Plant
- Geothermal Power Plant (GPP)
- Lignite Coal Power Plant
- LNG Power Plant
- Diesel Power Plant
- Naphtha Power Plant
- Wind Power Plant (WPP)
- Hard Coal Power Plant

4.1.1. Turkey's Electricity Production Forecast:

In our model, The Turkey electricity demand forecast is estimated by using the following variables through a multiple linear regression model. Generally, the model uses dummy variables for national and religious holidays and days of the week, and lagged values of Turkey's demand, and HDD and CDD values of certain metropolitan cities as input to make a demand forecast through a multiple linear regression model. The statistical model structure can be written in the following form:

$$\widehat{y_t} = \beta_0 + \sum_{i=1}^{7} \widehat{\theta_\iota} y_{t-i} + \sum_{i=0}^{81} \widehat{\beta_\iota} X_{i,t} + \sum_{i=0}^{81} \widehat{\gamma_\iota} Z_{i,t} + \sum_{i=0}^{12} \widehat{\delta_\iota} A_{i,t} + \sum_{i=0}^{7} \widehat{\omega_\iota} D_{i,t} + \sum_{i=0}^{n} \widehat{\varphi_\iota} T_{i,t} + \sum_{i=1}^{n} \widehat{\epsilon_\iota}\, \varepsilon_{t-i} \qquad (1)$$

In the above equation (1) $y_{t-i}, X_{i,t}, Z_{i,t}, A_{i,t}, T_{i,t}, D_{i,t}, \epsilon_{t-1}$ represent model explanatory variables, $\widehat{\theta_\iota}, \widehat{\beta_\iota}, \widehat{\gamma_\iota}, \widehat{\delta_\iota}, \widehat{\omega_\iota}, \widehat{\varphi_\iota}, \widehat{\epsilon_J}$, $(i = 0,1, \dots, n)$ represent the coefficients of the explanatory variables. Model inputs are:

Parameter	Description
β_0	Constant Value
y_t	Turkey's electricity demand at time t
$y_{t-\mathrm{i}}$	Turkey's electricity demand at time t-i (i days ago)
$X_{\mathrm{p},t}$	(Heating Degree Day) HDD value for province p at time t

$Z_{p,t}$	(Cooling Degree Day) CDD value for province p at time t
$A_{p,t}$	Dummy Variable for Months
$D_{\mathrm{p},t}$	Dummy variable for Days of the Week
$T_{p,t}$	Dummy Variable for Official Holidays and Similar Days

4.1.2. Run-of-river HPP Electricity Production Model

$$\widehat{y_t} = \beta_0 + \sum_{i=1}^{7} \widehat{\theta_\iota} y_{t-i} + \sum_{i=0}^{81} \widehat{\beta_\iota} X_{i,t} + \sum_{i=0}^{7} \widehat{\omega_\iota} D_{i,t} + \sum_{i=1}^{n} \widehat{\epsilon_\iota}\, \varepsilon_{t-i} \tag{2}$$

In the above equation (2) $y_{t-i}, X_{i,t}, D_{i,t}, \epsilon_{t-1}$ represent model explanatory variables, $\widehat{\theta_\iota}$, $\widehat{\beta_\iota}$, $\widehat{\omega_\iota}$, $\widehat{\epsilon_J}$, $(i = 0,1, \dots, n)$ represent the coefficients of the explanatory variables.

Model inputs (table):

Parameter	Description
β_0	Constant Value
y_t	Run-of-river electricity production at time t
$y_{t-\mathrm{i}}$	Run-of-river electricity production at time t-i (i days ago)
$X_{\mathrm{i},t}$	Rainfall amount for province i at time t
$D_{\mathrm{i},t}$	Dummy Variable for Days of the Week

4.1.3. Dam HPP Electricity Production Model

$$\widehat{y_t} = \beta_0 + \sum_{i=1}^{7} \widehat{\theta_\iota} y_{t-i} + \sum_{i=0}^{81} \widehat{\beta_\iota} X_{i,t} + \sum_{i=0}^{7} \widehat{\omega_\iota} D_{i,t} + \sum_{i=1}^{n} \widehat{\epsilon_\iota}\, \varepsilon_{t-i} \tag{3}$$

In the above equation (3) $y_t, X_{i,t}, D_{i,t}, \epsilon_{t-1}$ represent model explanatory variables, $\widehat{\theta_\iota}$, $\widehat{\beta_\iota}$, $\widehat{\omega_\iota}$, $\widehat{\epsilon_J}$, $(i = 0,1, \dots, n)$ represent the coefficients of the explanatory variables.

Model inputs (table):

Parameter	Description
β_0	Constant Value
y_t	Electricity production at time t from Dam HES
$y_{t-\mathrm{i}}$	Electricity production at time t-i (i days ago) from Dam HES
$X_{\mathrm{i},t}$	Rainfall amount for province i at time t
$D_{\mathrm{i},t}$	Dummy Variable for Days of the Week

4.1.4. Imported Coal Power Plant Electricity Production Model

$$\widehat{y_t} = \beta_0 + \sum_{i=1}^{7} \widehat{\theta_\iota} y_{t-i} + \widehat{\beta_\iota} X_{i,t} + \sum_{i=0}^{7} \widehat{\omega_\iota} D_{i,t} + \sum_{i=1}^{n} \widehat{\epsilon_\iota}\, \varepsilon_{t-i} \tag{4}$$

In the above equation (4) $y_t, X_{i,t}, D_{i,t}, \epsilon_{t-1}$ represent model explanatory variables, $\widehat{\theta_\iota}$, $\widehat{\beta_\iota}$, $\widehat{\omega_\iota}$, $\widehat{\epsilon_J}$ $(i = 0,1, \dots, n)$ represent the coefficients of the explanatory variables.

Model inputs:

Parameter	Description
β_0	Constant value
y_t	Imported coal production at time t
$y_{t-\mathrm{i}}$	Imported coal production at time t-i (i days ago)
$X_{\mathrm{i},t}$	Country-level electricity production at time t
$D_{\mathrm{i},t}$	Dummy Variable for Days of the Week

4.1.5. WPP Electricity Production Model

$$\widehat{y_t} = \beta_0 + \sum_{i=1}^{7} \widehat{\theta_\iota} y_{t-i} + \sum_{i=0}^{81} \widehat{\beta_\iota} X_{i,t} + \sum_{i=0}^{81} \widehat{\beta_\iota} {X_{i,t}}^2 + \sum_{i=0}^{81} \widehat{\beta_\iota} {X_{i,t}}^3 + \sum_{i=0}^{81} \widehat{\gamma_\iota} Z_{i,t} + \sum_{i=0}^{81} \widehat{\gamma_\iota} {Z_{i,t}}^2 + \sum_{i=0}^{81} \widehat{\gamma_\iota} {Z_{i,t}}^3 + \sum_{i=1}^{n} \widehat{\epsilon_\iota}\, \varepsilon_{t-i} \tag{5}$$

In the above equation (5) $y_t, X_{i,t}, Z_{i,t}, \epsilon_{t-1}$ represent model explanatory variables, $\widehat{\theta_\iota}$, $\widehat{\beta_\iota}$, $\widehat{\gamma_\iota}$, $\widehat{\epsilon_J}$ $(i = 0,1, \dots, n)$ represent the coefficients of the explanatory variables.

Model inputs:

Parameter	Description
β_0	Constant Value
y_t	RES Electricity Production at time t
$y_{t-\mathrm{i}}$	RES Electricity Production at time t-i (i days ago)
$X_{\mathrm{i},t}$	Maximum wind speed for province i at time t
$Z_{\mathrm{i},t}$	Average wind speed for province i at time t

4.1.6. Lignite Coal Power Plant Electricity Production Model

$$\widehat{y_t} = \beta_0 + \sum_{i=1}^{7} \widehat{\theta_\iota} y_{t-i} + \sum_{i=0}^{7} \widehat{\omega_\iota} D_{i,t} + \sum_{i=1}^{n} \widehat{\epsilon_\iota}\, \varepsilon_{t-i} \tag{6}$$

In the above equation (6) $y_t, X_{i,t}, D_{i,t}, \epsilon_{t-\mathrm{i}}$ represent model explanatory variables, and $\widehat{\theta_\iota}, \widehat{\beta_\iota}, \widehat{\omega_\iota}, \widehat{\epsilon_J}$ $(i = 0,1, \dots, n)$ represent the coefficients of the explanatory variables.

Model inputs:

Parameter	Description
β_0	Constant value
y_t	Lignite Coal Power Plant Electricity Production at time t
$y_{t-\mathrm{i}}$	Lignite Coal Power Plant Electricity Production at time t-i (i days ago)
$D_{\mathrm{i},t}$	Dummy Variable for Days of the Week

4.1.7. Other Resources Power Plant Electricity Production Model

For power plants that generate electricity from resources other than the ones for which production models were established above, our approach suggests that the production will be the same as the previous day. The model can be statistically expressed as follows:

$$\widehat{y_t} = y_{t-1} \tag{7}$$

Here, the estimated production value for day t is the production value for day t-1. The electricity production estimated with this model corresponds to the following supply sources and power plants: Asphaltite Coal Power Plant, Biomass Power Plant, Fuel Oil Power Plant, Solar Power Plant, Geothermal Power Plant, LNG Power Plant, Naphtha Power Plant, and Coal Power Plant.

These models collectively represent different methods for predicting electricity production based on a variety of factors, including the type of energy source used, historical production data, environmental factors like rainfall or wind speed, and even the day of the week.

Natural Gas Power Plants Electricity Production Forecast

We calculate the amount of electricity to be produced by Natural Gas Combined Cycle Power Plants according to the Merit Order approach by subtracting the production of other sources from Turkey's Electricity Production (1). It can be expressed through the equations numbered above as:

$$Natural\ Gas\ Combined\ Cycle\ Power\ Plants\ Electricity\ Production\ Forecast = 1 - (2 + 3 + 4 + 5 + 6 + 7) \qquad (8)$$

Natural Gas Power Plants Gas Consumption Forecast

The calculation of how much natural gas is consumed for the amount of electricity produced from Natural Gas Combined Cycle Power Plants on the relevant day is found by converting the unit of kWh to Sm3 and including the effect of the efficiency factor. The equation for this conversion calculation is presented below:

$$Natural\ Gas\ Combined\ Cycle\ Power\ Plants\ Natural\ Gas\ Consumption\ (Sm3) = kWh * \frac{10{,}64}{Percent\ of\ Productivity}$$

4.2. Residential Sector Natural Gas Consumption Model

The Province-Based Demand Forecast is estimated by our approach using the following variables through a linear regression model. In general, the model predicts demand using a multiple linear regression model, taking as input the instantaneous and delayed values of the HDD of the relevant province, the delayed values of the demand of the relevant province, and dummy variables for the days of the week. The model can be written in the following linear form.

$$\widehat{y_t} = \beta_0 + \sum_{i=1}^{7} \widehat{\theta_\iota} y_{t-i} + \sum_{i=0}^{3} \widehat{\beta_\iota} X_{t-i} + \widehat{\delta_\iota} Y_t + \sum_{i=0}^{7} \widehat{\omega_\iota} D_{i,t} + \sum_{i=1}^{n} \widehat{\epsilon_i}\, \varepsilon_{t-i} \qquad (9)$$

In the equation above $\widehat{\theta_\iota}$, $\widehat{\beta_\iota}$, $\widehat{\delta_\iota}$, ω_i, ϵ_i represent the coefficients of the model parameters, and $y_{t-i}, X_{t-i}, Y_t, D_{i,t}, \varepsilon_{t-i}$ $(i = 0,1, \dots, n)$ represent model explanatory variables.

Model inputs:

Parameter	Description
β_0	Constant Value
y_t	Demand for residential natural gas at time t on a province basis

$y_{t-\mathrm{i}}$	Demand for residential natural gas at time t-i (i days ago) on a province basis
X_t	HDD value for the relevant province at time t
X_{t-i}	HDD value for the relevant province at time t-i
$D_{\mathrm{i},t}$	Dummy variable for days of the week
Y_t	Dummy variable for the summer season

4.3. Turkey Natural Gas Consumption Model

The Turkey National Total Natural Gas Consumption Model is constructed with the aid of related parameters and total natural gas consumption data of the electricity and residential sectors. The Turkey Total Natural Gas Consumption Model is presented as follows:

$$\widehat{y_t} = \beta_0 + \sum_{i=1}^{7} \widehat{\theta_\iota} y_{t-i} + \widehat{\beta_\iota} X_t + \widehat{\gamma_\iota} E_t + \widehat{\delta_\iota} K_t + \sum_{i=0}^{7} \widehat{\omega_\iota} D_{i,t} + \sum_{i=1}^{n} \widehat{\epsilon_\iota}\, \varepsilon_{t-i} \qquad (10)$$

In the equation above $\widehat{\theta_\iota}, \widehat{\beta_\iota}, \widehat{\delta_\iota}, \omega_i, \epsilon_i$ represent the coefficients of the explanatory variables, and $y_{t-i}, X_t, E_t, K_t, D_{i,t}, \varepsilon_{t-i}$ $(i = 0,1, \ldots, n)$ represent model explanatory variables.

Model inputs:

Parameter	Description
β_0	Constant value
y_t	Total Natural Gas Demand in Turkey at time t
$y_{t-\mathrm{i}}$	Total Natural Gas Demand in Turkey at time t-i
X_t	HDD value for the relevant province at time t
E_t	Electricity Sector Natural Gas Consumption at time t
K_t	Residential Sector Natural Gas Consumption at time t
$D_{\mathrm{i},t}$	Dummy variable for days of the week

4.4. Industrial Sector Natural Gas Consumption Model

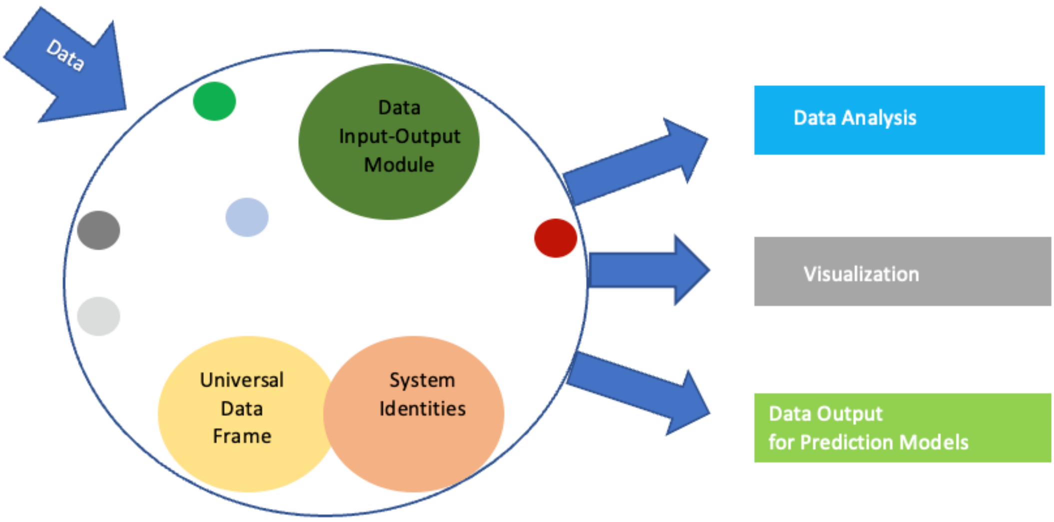
Since there is not a daily Industrial Sector Production Parameter, the Industrial Sector Natural Gas Consumption Model has not yet been established. The consumption model of this sector is formed by subtracting the Natural Gas Consumption Demand Estimates of the Electricity and Residential Sectors from the Turkey Total Natural Gas Demand Estimate. This model is formulated as follows:

$$\begin{aligned} &\mathit{Industrial\ Sector\ Natural\ Gas\ Demand} \\ &\qquad = \mathit{Turkey\ Total\ Natural\ Gas\ Demand} \\ &\qquad - \mathit{Electricity\ Sector\ Natural\ Gas\ Demand} \\ &\qquad - \mathit{Residential\ Sector\ Natural\ Gas\ Demand} \end{aligned}$$

5. PROPOSED DATA ANALYTICS FRAMEWORK

The proposed data management system aims to automate all processes of data pulling, classification, and formatting for analysis, improving the parts where the existing system is deficient and slow. Therefore, it is suggested that this system be written in Python due to the flexibility it provides in data input and output, as well as its speed, flexibility, and the robust tools and convenience it provides in data management and analysis. The data management system to be set up with Python will automatically pull the data from the daily report and classify it appropriately, and then this data will be easily transformed for analysis. Data can also be converted and stored in suitable formats. Since Python allows flexible data input and output, this system is expected to operate without any program data type restrictions. Additionally, with Python's data analysis tools, errors in the data can be detected and appropriately corrected. Moreover, with the analysis part working as an extension of this system, it will be possible to perform this error check even more robustly (Figure 2).

Figure 2. Proposed Data Management System



In the proposed system, all data should be kept in a hierarchical structure, and a system identity must be assigned for each output point. This system identity will be separate from the point code in the report, reflect all important characteristics of the point at first glance, and will also constitute the backbone of the hierarchical structure to be established. This identity will clearly determine which region and city the point is located in, which RMS-A point it is connected to if available, which RMS-A and RMS-B point it is in the city, and which consumer uses the natural gas coming out of this point. In this way, whether geographically, by point type, or by consumption type, these points will be arranged in a meaningful hierarchical order, greatly reducing the complexity of the system. Also, due to this hierarchical arrangement, the computational cost needed to create graphical analysis and different tables from the data will be reduced. The proposed identity system is shown in Table 2.

Table 2. Proposed Identity System

REGION CODE	CITY CODE	RMS-A CODE	RMS-B CODE	TYPE CODE
These are unique two-digit codes assigned to regions. They indicate the region in which the point is located.	The city plate code indicates the city in which the point is located.	These are three-digit codes representing the main exits in a city.	These are three-digit codes assigned to the secondary exits in a main exit.	These are three-letter codes that indicate the consumption category of the exit.

A system identity will be assigned to each output point in Turkey and this identity information will be kept in a separate structure. Historical consumption data will be matched with the relevant system identities and then a universal data frame will be created where all this data will be collected. Using this data frame and system codes, data can theoretically be displayed to include all desired information. For example, it will be possible to access all consumption data in Ankara province by using the region and license plate combination, and also to reach the consumption at a specific output point in Ankara province using the region, province, and the code given to that point. In addition, thanks to the general search characters (wildcard characters) that can be used in the search, more compact data frames for specific uses can be obtained from the universal data frame. For example, it will be possible to access natural gas consumption data used for industry between certain dates only in Ankara province. Thanks to this powerful and compact structure, data can also be easily classified and modeled for analysis and visualization.

6. CONCLUSION

Due to the rapidly increasing energy demand and a fifth of this demand being met by natural gas, natural gas consumption planning is of great importance. The most important input of this planning is natural gas consumption forecasts. In this article, the world literature on natural gas demand forecasting has been examined in detail, the current demand forecast situation in Turkey has been revealed, and a data analytics approach based on big data has been proposed. With the proposed approach whether geographically, by point type, or by consumption type, these points will be easily and meaningfully placed in a multiple hierarchical order, greatly reducing the complexity of the system. In addition, due to this hierarchical arrangement, the computational cost needed to create graphical analysis and different tables from the data will be reduced. Also, using the proposed data frame and system codes, data can theoretically be displayed to include all desired information. Thanks to the general search characters that can be used in the search, more compact data frames for specific uses can be obtained from the universal data frame.

REFERENCES

Aras, H., Aras, N. 2004 “Forecasting Residential Natural Gas Demand”, Energy Sources, 26(5), 463-472.

Aydin, G. 2015 “The Modeling and Projection of Primary Energy Consumption by The Sources”, Energy Sources A, 10(1), 67-74.

Azadeh, A., Zarrin, M., Beik, H.R. and Bioki, T.A. 2015 “A Neuro-Fuzzy Algorithm for Improved Gas Consumption Forecasting with Economic, Environmental And IT/IS Indicators”, Journal of Petroleum Science and Engineering, 133, 716-739.

Balestra, P. and Nerlove, M. 1966 “Pooling Cross Section and Time Series Data in The Estimation of a Dynamic Model: The Demand for Natural Gas”, Econometrica (pre-1986), 34(3): 585.

Baltagi, B. H., Bresson, G.and Pirotte, A. 2002 “Comparison of Forecast Performance for Homogeneous, Heterogeneous and Shrinkage Estimators: Some Empirical Evidence from US Electricity and Natural-Gas Consumption”, Economics Letters, 76(3): 375-382.

Beccali M, Cellura, M., Brano, V. L. and Marvuglia, A. 2004 “Forecasting Daily Urban Electric Load Profiles Using Artificial Neural Networks”, Energy Conversion And Management, 45(18-19): 2879-2900.

Berndt, E. R. and Watkins, G. C. 1977 “Demand for Natural Gas: Residential and Commercial Markets in Ontario and British Columbia”, Canadian Journal of Economics, 97-111.

Demirel, Ö. F., Zaim, S., Çalişkan, A. and Özuyar, P. 2012 “Forecasting Natural Gas Consumption in Istanbul Using Neural Networks and Multivariate Time Series Methods”, Turkish Journal of Electrical Engineering & Computer Sciences, 20(5): 695-711.

Dombaycı, Ö. A. 2010 “The Prediction of Heating Energy Consumption in A Model House by Using Artificial Neural Networks in Denizli–Turkey”, Advances in Engineering Software, 41(2): 141-147.

Durmayaz, A., Kadıoğlu, M. and Şen, Z. 2000 "An Application of The Degree-Hours Method to Estimate the Residential Heating Energy Requirement and Fuel Consumption in Istanbul", Energy, 25(12): 1245-1256.

Fagiani, M., Squartini, S., Gabrielli, L., Spinsante, S. and Piazza, F. 2015 "A Review of Datasets and Load Forecasting Techniques for Smart Natural Gas Water Grids: Analysis and Experiments", Neurocomputing, 170: 448-465.

Fan, Y. and Xia, Y. 2012 "Exploring Energy Consumption and Demand in China", Energy, 40: 23-30.

Görücü, F. B. 2004. "Evaluation And Forecasting of Gas Consumption by Statistical Analysis", Energy Sources, 26(3): 267-276.

Gracias, A.C., Lourenco, S.R. and Rafikov, M. 2012 "Estimation of Natural Gas Production, Import and Consumption in Brazil Based on Three Mathematical Models", Natural Resources, 3: 42-47.

Gutiérrez, R., Nafidi, A. and Sánchez, R. G. 2005 "Forecasting Total Natural-Gas Consumption In Spain By Using The Stochastic Gompertz Innovation Diffusion Model", Applied Energy, 80(2): 115-124.

Gümrah, F., Katircioglu, D., Aykan, Y., Okumus, S. and Kilincer, N. 2001 "Modeling of Gas Demand Using Degree-Day Concept: A Case Study for Ankara", Energy sources, 23(2): 101-114.

Ivezić, D. 2006 "Short-Term Natural Gas Consumption Forecast", FME Transactions, 34(3): 165-169.

Kaynar, O., Yilmaz, I. and Demirkoparan, F. 2010 "Forecasting of Natural Gas Consumption with Neural Network and Neuro Fuzzy System", In EGU General Assembly Conference Abstracts (Vol. 12, p. 7781).

Khan, MA. 2015 "Modeling and Forecasting the Demand for Natural Gas", Renewable and Sustainable Energy Reviews, 49: 1145-1159.

Kizilaslan, R. and Karlik, B. 2009 "Combination of Neural Networks Forecasters for Monthly Natural Gas Consumption Prediction", Neural network world, 19(2): 191.

Lim, H. L. and Brown, R. H. 2001. "Gas Load Forecasting Model Input Factor Identification Using a Genetic Algorithm". In Proceedings of the 44th IEEE 2001 Midwest Symposium on Circuits and Systems. MWSCAS 2001 (Cat. No. 01CH37257) (Vol. 2, pp. 670-673). IEEE.

Maddala, G. S., Trost, R. P., Li, H. and Joutz, F. 1997 "Estimation of Short-Run and Long-Run Elasticities of Energy Demand from Panel Data Using Shrinkage Estimators", Journal of Business & Economic Statistics, 15(1): 90-100.

Merkel, G., Povinelli, R. and Brown, R. 2018 "Short-Term Load Forecasting of Natural Gas with Deep Neural Network Regression", Energies, 11(8): 2008.

Panella, M., Barcellona, F. and D'Ecclessia, R.L. 2012 "Forecasting Energy Commodity Prices Using Neural Networks", Research Article. DOI:10.1115/2012/289810.

Pang, B. 2012 "The Impact of Additional Weather Inputs on Gas Load Forecasting".

Pappas, S. S., Ekonomou, L., Karamousantas, D. C., Chatzarakis, G. E., Katsikas, S. K. and Liatsis, P. 2008 "Electricity Demand Loads Modelling Using Autoregressive Moving Average (ARMA) Models", Energy, 33(9): 1353-1360.

Payne, J. E., Loomis, D. G. and Wilson, R. 2011 "Residential Natural Gas Demand in Illinois: Evidence from The ARDL Bounds Testing Approach". Journal of Regional Analysis and Policy, 41(1100-2016-89863), 138.

Potočnik, P., Thaler, M., Govekar, E., Grabec, I. and Poredoš, A. 2007 "Forecasting Risks Of Natural Gas Consumption In Slovenia". Energy Policy, 35(8): 4271-4282.

Sarak, H. and Satman, A. 2003 "The Degree-Day Method to Estimate the Residential Heating Natural Gas Consumption in Turkey: A Case Study", Energy, 28(9): 929-939.

Soldo, B. 2012 "Forecasting Natural Gas Consumption", Applied Energy, 92: 26-37.

Suganthi, L., and Samuel, A.A. 2012 "Energy Models for Demand Forecasting-A Review", Renewable and Sustainable Energy Reviews, 16:1223-1240.

Szoplik, J. 2015 "Forecasting of Natural Gas Consumption with Artificial Neural Networks", Energy, 85: 208-220.

Tamba, J. G., Essiane, S. N., Sapnken, E. F., Koffi, F. D., Nsouandélé, J. L., Soldo and B., Njomo, D. 2018 "Forecasting Natural Gas: A Literature Survey", International Journal of Energy Economics and Policy, 8(3): 216-249.

Taşpınar, F., Celebi, N. and Tutkun, N. 2013 "Forecasting of Daily Natural Gas Consumption on Regional Basis in Turkey Using Various Computational Methods", Energy and Buildings, 56: 23-31.

Vitullo, S., Brown, R. H., Corliss, G. F. andMarx, B. M. 2009 "Mathematical Models For Natural Gas Forecasting", Canadian Applied Mathematics Quarterly.

Voudouris, V., Matsumoto, K., Sedgwick, J., Rigby, R., Stasinopoulos, D. and Jefferson, M. 2014 "Exploring the Production of Natural Gas Through the Lenses of The ACEGES Model", Energy Policy, 64: 124-133.

Xiong, P.P., Dang, Y.G., Yao, T.X. and Wang, Z.X. 2014 "Optimal Modeling and Forecasting of The Energy Consumption and Production In China", Energy, 77: 623-634.

Hakem Değerlendirmesi: Dış bağımsız.

Çıkar Çatışması: Yazar çıkar çatışması bildirmemiştir.

Finansal Destek: Yazar bu çalışma için finansal destek almadığını beyan etmiştir.

Teşekkür: -

Peer-review: Externally peer-reviewed.

Conflict of Interest: The author has no conflict of interest to declare.

Grant Support: The author declared that this study has received no financial support.

Acknowledgement: -