SPE 91413

Forecasting US Natural Gas Production into year 2020: a comparative study.

Anangela Garcia, SPE, Dominion Resources and Shahab D. Mohaghegh, SPE, West Virginia University

This paper was prepared for presentation at the SPE Eastern Regional Conference and Exhibition held in Charleston, West Virginia, 15-17 September 2004.

Summary

There are several models for predicting US natural gas production. The most referenced model is the one developed and used by the Energy Information Administration (EIA). Recently a model was developed at Texas A & M University that used neural networks in order to learn from the past when predicting the future. One of the major short comings of the existing models is the fact that they appear to be too certain about their predictions. In other words, they fail to take into account the uncertainties associated with predicting many of the parameters that are inputs into their models and treat them as a certain single values.

In this paper authors present a new model with several advantages over the existing models. First, this model treats the US Natural Gas Production as a time series with all its inherent complexities and incorporates the required methodologies during the development process. Second, this model does not use parameters such as gas price as an input. If we know the future price of gas, then predicting the production levels will not be so difficult. Third, this model does not assume that crisp and exact information is available on the future values of all the input parameters such as GDP, Population and average depth of oil and gas wells. Instead by incorporating a Monte Carlo simulation approach, it uses a probability distribution function for each of its input parameters, and hence the outcome of the model (US Natural Gas Production) is not a crisp number for each year rather it is a range that includes a minimum, a maximum and a most likely value.

The model is developed using the state-of-the-art in intelligent systems by using data up to year 1998. The model is then validated by comparing its results with actual US Natural Gas production for years 1998-2002 as blind (verification) years. During this five-year period this model outperforms all the existing models in the literature. The US Natural Gas production is then forecasted until year 2020.

Introduction

Natural gas has become one of the most important sources of energy in the United State during the past decade. According to the Energy Information Administration (EIA), the consumption of natural gas in USA rose from 10.24 TCF in 1960 to 17.37 TCF in 2001. It is apparent that natural gas consumption will increase in the future. For 2002, the demand of gas was 22.52 TCF, and 100% of this gas was consumed in the United States. The proven reserve in the United States was estimated to be 183.43 TCF in 2001. The major consumers of natural gas in the U.S. are the industrial sector (35%), followed by the residential (25%), electric energy (24%), and the commercial sector (16%)[1].

The increasing demand of natural gas has forced the oil and gas industry to constantly seek new approaches and efficient methods to enhance the national gas supply. There is also an intensive ongoing effort for developing new methodologies for predicting the future natural gas demand and production. Identifying the variables that influence the consumption, demand and production of natural gas is a complex and non-trivial task. Many important decisions, risk evaluations, and economic analysis are made based on these predictive models. It is notable that the predictive models use input variables that themselves need to be predicted or at least guessed in order to be use in the model. This makes the task of building a predictive model for US natural gas production quite challenging.

Currently, there are several models for predicting US natural gas production. The most referred model is the one developed and used by the Energy Information Administration. EIA prepares the Annual Energy Outlook, which presents a midterm forecast and analysis of US production, supply, demand, and prices through 2025. The projections are based on results from EIA's National Energy Modeling System (NEMS)[1]. Recently a new model was developed at Texas A&M University[2] that introduced a new way of modeling the US Natural Gas production. The Texas A&M model used artificial neural networks in order to learn from the past while predicting the events in the future. Al-Fattah and Startzman used a backpropagation, three layer neural network in order to forecast US natural gas supply to the year 2020. The results indicated that the US would maintain its 1999 production of natural gas to 2001. The neural network showed that natural gas production would increase during the period 2002 to 2012 on an average rate of 0.5% per year. This increased rate will be more than double for the period 2013-2020[2].

The Al-Fattah and Startzman's neural network forecast was based on several assumptions such as:

- the gas prices were expected to increase by 1.5% per year.
- the gas depletion rate were expected to increase by 1.45 per year.
- drilling of gas exploratory wells would improve by 3.5% per year.
- drilling of oil/gas exploratory wells will increase an average of 2.5% per year.
- GDP would have an average increase of 2.1% per year.

The inputs used in the Texas A&M model were number of gas exploratory wells, oil/gas completed exploratory wells, oil exploratory wells, gas depletion rate, depth drilled per wells in gas/oil exploratory wells, footage drilled in gas exploratory wells, proved reserves, gas wellhead prices, and growth rate of the growth domestic product. Al-Fattah and Startzman's approach was a step forward in developing predictive models for commodities that appear to have unconventional and sometime chaotic behavior. The behavior is referred to as chaotic since many uncontrollable events play important roles in shaping it.

Authors feel that Texas A&M model can be improved by changing two important factors. First, the gas price should be removed from the list of input parameters. It may be argued that predicting wellhead gas price is as important and challenging as predicting the US natural gas production and that if one is known the other can be estimated with reasonable degree of accuracy. Another argument may be that the same approach that is used to estimate the future gas price so it can be used as an input to the neural model may be equally effective for forecasting the US natural gas production. On the other hand, removing such an important and influential parameter from the list of inputs to the model will make the process of neural model building quite challenging but nevertheless, if successful, much more reliable.

The second improvement to Texas A&M model can be in the form of incorporation of the uncertainties that is inherent in the input parameters used in the neural model. For this, authors recommend the use of Monte Carlo simulation. In this paper authors present a new forecasting model using the state-of-the-art in intelligent systems. The proposed methodology takes into account the uncertainty associated with the input parameters that influence the gas production supply excluding wellhead gas price and some other parameters.

Methodology

The first stage of this study consisted of an intensive research for information and available data. Two of the major sources of data were the Energy Information Administration and the American Petroleum Institute[5]. Data from 1947 to 2002 was used for train, calibrate and verify the neural network. Data such as, US crude oil reserves, US total wells reported as completed, US exploratory wells and footage, the supply of US dry natural gas, estimated natural gas reserves, US natural gas production, gross domestic product, population and some other parameters were used as a starting point for analyzing and evaluating the collected data trends. All the variables were tabulated and normalized in order to standardize the values of each parameter.

For the second stage, a large number of plots were constructed based on all the possible combinations between the variables. The relationship between the variables were studied and evaluated in order to identify any apparent trends. Evaluation of parameters' influence on the US natural gas production was performed using fuzzy combinatorial analysis[3]. This step played an important role in identifying the parameters that were used as the input variables into the final neural model. Upon completion of this step following parameters were used as input to the neural network model:

1. US natural gas production from previous year.
2. Growth Domestic Product.
3. US Population.
4. Average Depth of Oil & Gas Exploratory Wells.
5. Annual Gas Depletion Rate.

The natural gas production forecast is based on historic data from 1947 to 2002 and the US natural gas production is estimated for the next 17 years (2003-2020). US natural gas production is essentially a time-series. The type of neural networks that have proven to be effective in predicting time-series are a family of networks called recurrent neural networks. Recurrent networks provide a feedback loop either from the hidden layer into the input layer or from the output layer into the input layer. As shown above, in the model presented in this paper the feedback from the output layer into the input layer has been implemented.

Furthermore, fuzzy cluster analysis was used in order to guide the partitioning of the training and calibration data sets. This is a unique feature of the neural network package[4] that was used during the development of this model. This process, along with the fuzzy combinatorial analysis for input parameter identification markedly increases the chances of developing successful neural models.

Once the neural model is built, calibrated and verified using blind data, it needs to be used in its predictive mode. This simply means that input values are provided to the network and it will forecast the US natural gas production. The model's prediction is a resultant of the input data and any assumption made to generate those values will influence the forecast. Therefore, using crisp values for the input variables would result in risky predictions.

In order to address this concern a Probability Distribution Function (PDF) was proposed for each of the input parameters instead of using single values as input to the model. This task was accomplished by identifying the trend of each variable versus time. Three different scenarios where proposed in order to handle the uncertainty of estimating the future input values. Maximum, most likely and minimum trends were established for each input.

Figures 1 through 4 show the potential trends that were identified for each of the inputs. It is important to mention that these trends were product of the authors' judgments and one may choose to move the minimum, maximum or most likely trends per one's view of the future trends. Each author performed this procedure independently, and then the results were combined in order to generate two different versions (scenarios) of the same graph for each input. However the trends were not so different from one another. The point is being made that it is more likely to agree on such trends than to agree on certain crisp percentages of change for a particular parameter.

In order to forecast the US natural gas production for any particular year, 500 neural model runs were made. Following a Monte Carlo simulation methodology, each neural model run was performed by identifying a randomly selected value for each of the input parameters based on their PDF. The results of the 500 runs were then plotted in order to identify the probability distribution function of the US natural gas production for that particular year. This resulting PDF would then be used as the PDF for the Monte Carlo simulation of the next year.

Results and Discussion

As was mentioned before, the model was trained and calibrated (tested) using data from 1947 to 1997. Then the trained model was applied to predict the US natural gas production from 1998 to 2020. The US natural gas production from 1998 to 2002 was used as the blind (verification) dataset to check the integrity and the ability of the model to forecast US natural gas production into the future since the actual values of these years were available.

Figure 5 shows the history of the US natural gas production along with the forecast until the year 2020 as modeled by previously developed models as well as the model developed in this study. The stochastic model as well as the EIA model forecast an ever growing production with time. These models seem to honor the production trend demonstrated by early years, although the EIA model tends to forecast a slow down as we approach year 2020. The Texas A&M model and the WVU model seem to agree the the rate of growth in production is much slower than that predicted by other two models. This can be attributed to the fact that Texas A&M model and WVU model have incorporated the previous data in a different fashion using an intelligent systems approach as compared to other two models.

Figure 6 shows only the historical US natural gas production along with the forecast from the WVU model. The production values from the years that have been used as verification are identified in this figure. It can be observed that the actual productions and the forecasted results from the WVU model for these five years (1998-2002) are in good agreement. Years 1998 and 1999 match with the actual values almost perfectly. For year 2000, the predicted value is slightly overestimated, but it is still quite close to the actual production. In 2001 and 2002, the actual values fall into the predicted range generated by the WVU model. The expected natural gas production is presented as a range that includes maximum, most likely and minimum for production from years 2003-2020.

Figure 7 is a closer look at the four US natural gas production forecasting models namely the EIA model, the Stochastic model, the Texas A&M model and the WVU model. This figure allows a closer inspection of the forcast performed by each of these models.

WVU model is the only model that incorporates the uncertainty associated with the input parameters and provides user with a range of possible production values rather than a single value. It is also the only model (to the best of our knowledge) that uses a feedback technique for predicting the production of each year. It uses the previous year's production as an input for the present year. Therefore, it is recommended to use the data that becomes available each year and retrain the network. This results in an organic model that evolves in its forecast accuracy with time and become more and more reliable.

WVU model seems to predict a higher production in early years as compared to Texas A&M model and as data from these years becomes available and the model is retrained with newer data its forecast might be modified. We plan to train and maintain an updated model on our department's web site for public use.

Conclusions

A gas-forecasting model was developed to predict the US natural gas production until 2020. The WVU model was built based on a recurrent neural network in order to accommodate the time series nature of the US natural gas production. Different tools were used in order to build the WVU Model. Fuzzy combinatorial analysis was applied to identify the most influencial parameters for predicting the US natural gas production. Fuzzy cluster analysis was used to guide the partitioning of the data into the training and calibration sets. This played an important role in developing a successful neural network. The degree of uncertainty of the inputs was addressed by using Monte Carlo Simultion through defining probability distribution functions for each of the input parameters.

References

1. Energy Information Administration web site. http://www.eia.doe.gov
2. Al-Fattah S. M. and Startzman R. A.: "*Predicting Natural Gas Production using artificial neural network*". SPE 68593 (April, 2001).
3. Mohaghegh, S. D. "Essential Components of an Integrated Data Mining Tool for the Oil & Gas Industry, With an Example Application in the DJ Basin", SPE 84441, Proceedings, 2003 SPE Annual Conference and Exhibition, October 4-October 8, Denver, Colorado
4. IDEA™ from Intelligent Solutions, Inc. http://www.intelligentsolutionsinc.com/idea.htm
5. American Petroleum Institute: "*Basic Petroleum Data Book, Petroleum Industry Statistics*". Volume XXI, Number 2, Washington, DC (2001).
6. Mohaghegh, S.D. "Virtual Intelligence Applications in Petroleum Engineering: Part 1 ; Artificial Neural Networks", Journal of Petroleum Technology, Distinguished Author Series, September 2000, pp64-73.

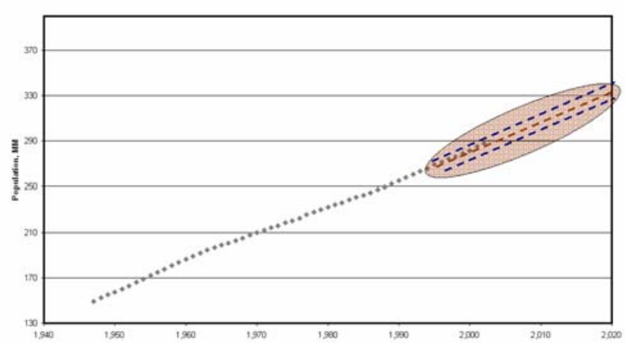

Figure 1. Estimated Population potential trend (maximum, most likely, and minimum) for 2003 to 2020.

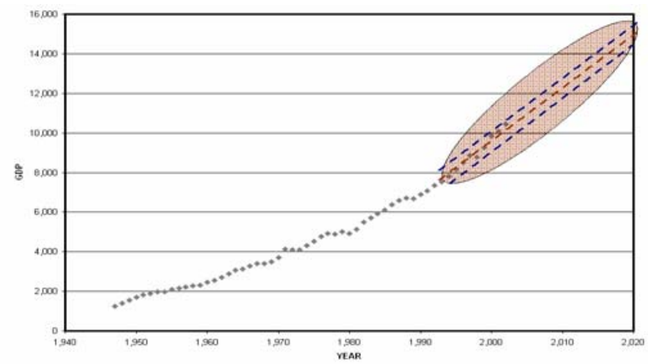

Figure 2. Estimated Gross Domestic Product potential trend (maximum, most likely, and minimum) for 2003 to 2020.

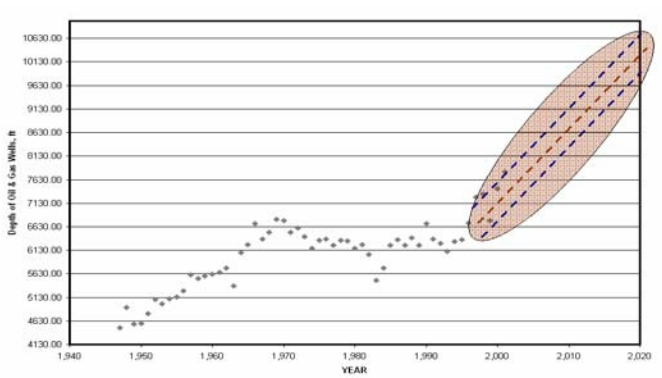

Figure 3. Estimated Average Depth of Oil-Gas exploratory wells potential trend (maximum, most likely, and minimum) for 2003 to 2020.

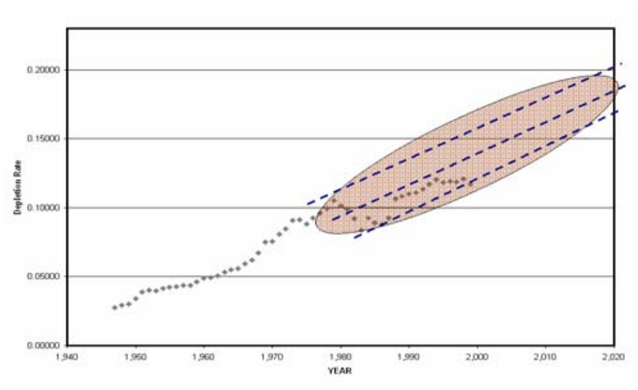

Figure 4. Estimated Annual Gas Depletion rate potential trend (maximum, most likely, and minimum) for 2003 to 2020.

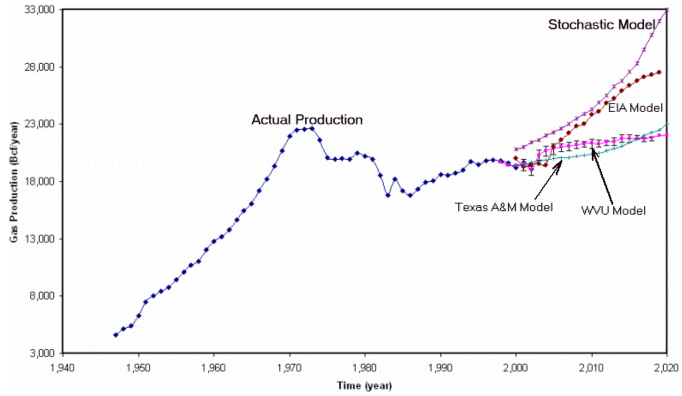

Figure 5. US Natural Gas forecasting from 2003 to 2020, using all available models.

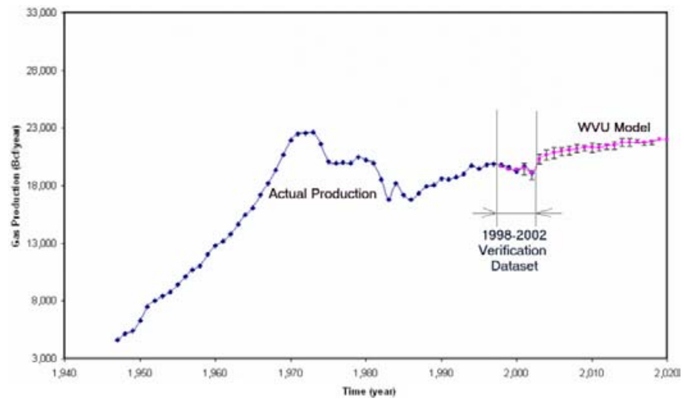

Figure 6. US Natural Gas forecasting from 2003 to 2020, using the WVU model.

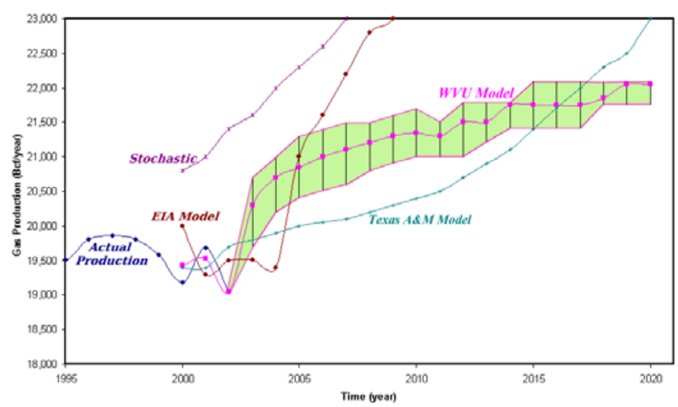

Figure 5. Forecasting comparison of five different models for prediction of US Natural Gas forecasts from 2003 to 2020.