Open Access Journal

Journal of Power Technologies 99 (4) (2019) 272–280

journal homepage:papers.itc.pw.edu.pl

Long-term prediction of underground gas storage user gas flow nominations

Teresa Kurek*, Konrad Wojdan, Konrad Świrski

Institute of Heat Engineering, Faculty of Power and Aeronautical Engineering, Warsaw University of Technology, Nowowiejska 21/25, 00-665 Warsaw, Poland

Abstract

Many companies operating on the natural gas market use natural gas storage to balance production and transport capacities with major variations in gas demand. This paper presents an approach to predicting users' gas flow nomination in underground gas storage by different users. A one-year prediction horizon is considered with weekly data resolution. Basic models show that whereas for the great majority of users we can predict nomination based only on weather data and technical parameters, for some users additional macro-economic data significantly improved prediction accuracy. Various modeling techniques such as linear regression, autoregressive exogenous model and Artificial Neural Network were used to develop prediction models. Results show that for most users an Artificial Neural Network provides optimal accuracy, indicating the non-linearity of the relationship between input and output variables. The models developed are intended to be used as support for facility operation decisions and gas storage product portfolio modifications.

Keywords: underground gas storage; long term forecasting; artificial neural networks; prediction models; demand forecasting

1. Introduction

Natural gas is produced in roughly constant quantities throughout the year. A large part of gas supplies are transported immense distances. Gas consumption by industry, power stations and households varies according to the season and time of day. Demand is much higher in winter than in summer, and more gas is used during the day than at night. To adjust to the seasonal variations in demand and the daily peaks, gas must be held in underground gas storage (UGS) facilities. UGS are described in detail in [1].

Much research has been done on UGS. Some researchers focus on directions in UGS development such as steady gas supply, energy balancing and inflow performance [2, 3, 4]. Together with the development of UGS facilities, research has focused on UGS optimization, in both economic [5] and technical terms [6, 7]. UGS facilities can be optimized through improving gas compressor performance [7] and/or gas flow distribution [6]. Another approach to UGS optimization is to define an optimal strategy (decision rules) for plant operation.

As presented in [7] aggregated predicted gas flow nominations for storage services is a main optimization algorithm input variable. Storage services demand is expressed in hourly nominations, i.e., hourly values of gas flow and flow direction to/from UGS (negative nominations indicate gas withdrawal from UGS and positive nominations indicate gas injection). Usually, gas flow nomination trajectory is known one day in advance. UGS users are obligated to submit short-term hourly gas flow nomination schedules. That way, UGS operators can safely prepare the facility for nominated gas flow injections/withdrawals. However, to optimize UGS operations an extended horizon of defined gas flow nomination should be considered, in particular a one year horizon is recommended, taking into account the one year UGS working cycle. A long-term gas flow nomination horizon is required to best optimize storage. Technical conditions affect the storage caverns/wells (depending on UGS type), which depend on the storage filling level. Therefore short-term operational decisions impact UGS working conditions over the one year time horizon.

Prediction models for UGS are made mainly for estimating cavern depth (e.g. pre-construction, feasibility study) or injection performance. Optimal cavern depth, production and injection performance are estimated in [8] based on IMEX simulation. In [9] the authors present a downhole inflow performance forecast based on data from field development used to predict inflow performance. The prediction was made for 15 injection-production cycles and appeared to be

*Corresponding author

Email address: teresa.bil@itc.pw.edu.pl (Teresa Kurek)

better than the traditional equation.

The natural gas market is strongly connected to UGS work (e.g., users can buy a larger amount of gas—if the price is lower—and store it in UGS) therefore gas demand forecasting has been an area of interest in many papers. In [10] the authors present models for forecasting residential monthly natural gas consumption. The method relies on dividing a year into two seasons—heating and non-heating—and estimating individual autoregressive time series models for each period. For Ankara, Turkey an Artificial Neural Network based model for forecasting gas consumption [11] and model using the degree-day concept is presented in [12]. Prediction models forecasting industrial end-use natural gas consumption in a 1–3 year horizon are presented in [13]. The forecast is obtained by combining three different components: one that captures the trend of the time series, a seasonal component based on the Linear Hinges Model, and a transitory component to estimate daily variations using explanatory variables. Gas demand forecasting models have been developed for many countries e.g. Spain [14], Turkey [10], India [15] Taiwan [16].

The papers above present gas demand forecasting, which plays an important role in most decisions in the natural gas market (especially for entities involved in trade) and transmission grid management. In the case of UGS management, general gas demand forecasting is insufficient. There is a need to develop an algorithm to predict long-term demand for storage services. Storage services demand is also an input variable for most of the UGS optimization algorithms described earlier.

The main interest in this paper is to develop models that are able to predict the gas flow nomination of UGS users in a one year horizon with weekly resolution. Nominations defining demand for gas storage services are expressed in kWh. A negative nomination signifies gas withdrawal from UGS and a positive nomination signifies gas injection to storage.

The outline of this paper is as follows: Historical data used to develop models is introduced in section 1. Data analysis is presented in section 2. Modeling techniques are presented in section 3. Section 4 describes the process of modeling and gives the results obtained.

2. Theory

2.1. Historical Data

The models are developed based on historical data from a real underground gas storage location in Europe. In the UGS presented in this paper, the gas is stored in porous structures of former natural gas reservoirs. Gas is injected into storage via dedicated wells with compressors.

In this paper four selected UGS user profiles are presented that store gas in the analyzed UGS. Prognostic models of those four users were developed on historical data from a period of 3 years (2013–2016), except for one user where the historical data contained measurements from one year (July 2014–June 2015)

Table 1: UGS user limitations

Factor	Description	Unit
LIMIT_ENTRY_MAX_QN	Maximum rate for gas injection	kWh
LIMIT_EXIT_MAX_QN	Maximum rate for withdrawal	kWh

Table 2: Weather data

Variable	Description	Unit
TEMP	Daily average temperature	Fahrenheit
TEMP_MAX	Maximum temperature	Fahrenheit
TEMP_MIN	Minimum temperature	Fahrenheit
WIND_SPEED	Daily average wind speed	knots
MAX_WIND_SPEED	Maximum wind speed	knots

For legal reasons, users' gas flow nominations were normalized; this operation has no impact on the performance of models presented in this paper (general behavior and nomination trends were preserved).

2.1.1. Models output

Models were developed to predict gas storage users' gas flow nominations with a one year prediction horizon (weekly aggregated). Historical data contains a daily nomination, therefore weekly aggregation of nominations were calculated according to formulas:

$$w_{compensation} = \frac{7}{n_{week\ days}} \tag{1}$$

$$y_{agg} = w_{compensation} \cdot \sum_{day=1}^{7} y_{day} \tag{2}$$

where: y_{agg}—aggregated value of gas flow nominations; $n_{week\ days}$—number of days in week; y_{day}—day value of gas flow nomination; $w_{compensation}$—factor was added to overcome the issue of weeks where the number of days is lower than 7 (for example, the first week of the year starting on a Tuesday) therefore the aggregated value for that week would be much lower than for full weeks.

2.1.2. Limits and working gas volume

Each storage user determined (signed in its contract) its own storage capacity factors, as presented in Table 1.

Limit variables indicate that non-linear models can give better prediction accuracy than linear models.

2.1.3. Weather data

Weather data was downloaded from the National Center for Environmental Information [17] and included data from weather stations in the capital cities of European Union member states and Switzerland. The data for each weather station comprised daily measurements from January 2013 to December 2015 and is presented in Table 2. Weather data was chosen as a model input variable, because UGS users are trading in the whole of Europe and weather data is the main input in gas demand forecasting models ([10, 11, 12, 13, 14, 15, 16]).

Table 3: Additional weather data

Variable	Description
WIND_SP_SQRT	Square root of daily average wind speed
WIND_SP_SQRT *TEMP	Product of square root of daily average wind speed and daily average temperature

Table 4: Gas prices

Variable	Description	Units
Germany	Russian Natural Gas border price in Germany	$\frac{USD}{thous.\ m^3}$
Henry_Hub	Natural Gas spot price at the Henry Hub terminal in Louisiana	$\frac{USD}{thous.\ m^3}$
Japan	Indonesian Liquefied Natural Gas in Japan	$\frac{USD}{thous.\ m^3}$

Weather variables used as model inputs were calculated as a weighted average according to formulas:

$$TEMP = \frac{\sum_{country}^{j=1} w_j \cdot TEMP_j}{\sum w_j} \tag{3}$$

$$WINDSPEED = \frac{\sum_{country}^{j=1} w_j \cdot WINDSPEED_j}{\sum w_j} \tag{4}$$

where: TEMP—weight average of a temperature; $TEMP_j$—temperature in capital of j country; w_j—population of j country; WIND SPEED—weight average of a wind speed; $WIND\ SPEED_j$—wind speed in j country.

Finally, additional variables were added to the set of possible input variables. Variables are presented in Table 3.

2.1.4. Gas prices

Gas prices contain the data set out in Table 4 with monthly granulation. The data was extended to daily granulation by linear interpolation. Despite the fact that prediction is made for gas nominations with one week granulation, gas prices data was extended to daily granulation to create more input variables (different data delays) and produce better correlation analysis; the correlations coefficients were calculated for daily delayed data.

Table 5: Macro-economic data

Variable	Description	unit
DGP	Domestic Gross Product for Germany	millions
EV_PETROL	Evolution Petroleum Corporation stock prices for market opening	USD
Ind_Prod	Industrial Production for Germany	-
Ind_Prod_IDX	Industrial Production Index for Germany	-
Oli_IL	Historical prices for Oil India Limited	USD
CO2_emission	CO2 emission price on the European Climate Exchange	$\frac{EUR}{ton_{CO2}}$
CO2_open	CO2 emission price on the European Climate Exchange, price for market opening	$\frac{EUR}{ton_{CO2}}$
CO2_high	CO2 emission price on the European Climate Exchange the highest price of the day	$\frac{EUR}{ton_{CO2}}$
CO2_low	CO2 emission price on the European Climate Exchange, the lowest price of the day	$\frac{EUR}{ton_{CO2}}$

Table 6: Additional data

Variables	Description
JANUARY FEBRUARY MARCH APRIL MAY JUN JULY AUGUST SEPTEMBER OCTOBER NOVEMBER DECEMBER	1—for days in specified month 0—for days not in specified month
SW_Christmas	1—for Christmas days 0—rest days
SW_Easter	1—for Easter days 0—rest days
SW_1May	1—1st of May 0—rest days
SW_New_year	1—for New Year's Eve and New Year's Day 0—rest days
SW_1Nov	1—First November 0—rest days
SW_15Aug	1—15th August 0—rest days
SW_6Jan	1—6th January 0—rest days
SW_CCh	1—Corpus Christi 0—rest days
week_nb	Week number (1,2,3...52)
month_nb	Month Number (1,2,3,...,12)

2.1.5. Macro-economic data

Due to the varied uses of natural gas, such as electricity production, industrial and residential use, some macroeconomic factors were considered as possible input variables. The factors are presented in Table 5; for each factor where values granulation is less dense than daily values, the data was extended to daily granulation by linear interpolation. The data were extended to daily granulation for the same reasons as the gas prices data.

2.1.6. Defined data

User gas flow nomination signs (positive for injection, negative for withdrawal) for certain seasons are repeatable (gas injection to UGS during summer and gas withdrawal from UGS during winter) therefore some additional variables were created. Newly-created variables describe the current season of the year. Additional variables describing periods of international holidays were created, since those days can be the cause of unusual user behavior. The newly-created variables are presented in Table 6.

If in a given week there are the last days of April and the first days of May then variables: APRIL, MAY and SW_1May have the value 1.

2.2. Data delays

Variables presented in sections 1.2–1.6 were considered as an input variable set from which input data for each model was selected. This variable set was extended by additional variables created by shifting weather data and gas prices data. The weather data was delayed daily—the maximum delay was 31 days—and accelerated daily: the maximum acceleration was 7 days. Gas prices were only delayed with a maximum delay of 31 days. Daily shifting data was considered to multiply and diverse possible input variables (user can rely on historical weather data while making decisions about gas flow nominations), also daily modeling is more efficient than weekly (during aggregation we lose information). Furthermore, weather data was accelerated in light of UGS users' operating activities: since gas demand is dependent on temperature, users can rely on weather forecasting (in our case historical data acceleration) to determine gas flow nomination for the immediate future period (e.g. week).

Variable delay is marked at variable description with the signs (-) for delay and (+) for accelerate. After the sign, the number of days for shifting is noted, for example feature TEMP-4 means that feature TEMP has been delayed for four days. Data weekly aggregation was performed after delaying the daily data. Gas flow nominations were also delayed daily (to analyze autocorrelation) with a maximum delay of 365 days (one year). In ARX models gas flow nominations delayed for seven days (one week) were used due to the output data resolution requirement being one week.

2.3. Data analysis

Historical data presented in section 2 was analyzed to discover main tendencies and behavior of users.

Several main conclusions were obtained:

1. Mean temperatures and max temperatures have bimodal distribution, and minimum temperatures have unimodal distribution with a skewing tendency to the right side of the histogram. Histograms representing distributions are presented in Fig. 1.
2. All UGS users have a general annual trend: withdrawal in winter and injection in summer.
3. User nomination trends vary from user to user, showing different user profiles. User profiles are dependent on user operation areas, like industry, local market etc.
4. For some users we can see one week autocorrelation up to 5 weeks in the past. In Fig. 2 the autocorrelation coefficient (Pearson coefficient) values through the year are presented for one user. The upper graph shows a one year horizon and the bottom graph shows a three month horizon.
5. For all users we can see high correlation with temperature (mean, maximum and minimum), moreover for one user gas flow nomination correlation with LIMIT_EXIT MAX_QN is 0.31 (for entry, the correlation is 0.21) for another user the correlation with a limit both for exit and entry is 0.15, for all other users the correlation with injection and withdrawal limits (Pearson correlation coefficient) is below 0.07.
6. For most input variables the Pearson and Spearman correlation coefficients have a similar value.

2.4. Modeling

Underground gas storage user behavior varies depending on user profile. In this paper four users of UGS are presented. For each user various prediction techniques were tested to obtain the best prediction model. As is presented, for each user a different technique gave the best solution. The reason for this is that users use gas storage for different purposes. A short summary of the modeling techniques used is presented below.

Table 7: Abbreviations

Abbreviation	Model
LB	Linear Basic
LA	Linear Advanced
LBA	Linear Basic with autoregression
LAA	Linear Advanced with autoregression
ANNB	Artificial Neural Network Basic
ANNBA	Artificial Neural Network Basic with Autoregression
ANNA	Artificial Neural Network Advanced
ANNAA	Artificial Neural Network Advanced with Autoregression

2.4.1. Linear Regression

Linear regression (LR) assumes that the output value is the sum of input variables multiplied by determined weight plus intercept. The main assumption of linear regression theory is that input and output variables are linearly dependent on each other. The unknown intercept and weight are determined through the least squares method. Input variables were selected by forward feature selection.

2.4.2. Autoregression

Autoregression with exogenous inputs (ARX) is similar to linear regression models, the only difference being that historical prediction values are considered as additional input. Model can be represented algebraically as:

$$y_t = F\left(y_{t-7}, u_t, u_{t-1}, ...,\right) \quad (5)$$

where: y_t—prediction for time t, $y_{t\text{-}7}$—historical prediction from one week before, u_t—external variables. F is approximation function. In this paper linear regression and an Artificial Neural Network were used as approximation functions. External variables were selected from historical data presented in section 1.

2.4.3. Artificial Neural Network

Artificial Neural Network (ANN) is a non-linear approximation method. The advantage of using ANN is the good modeling it gives of non-linear dependencies, like in the case presented in the paper: LIMIT_ENTRY_MAX_QN and LIMIT_EXIT_MAX_QN. Input variables were selected based on forward feature selection results, Pearson and Spearman correlation analysis and personal conclusions based on knowledge of UGS operation.

3. Results and discussion

In this section the abbreviations presented in Table 7 are used. Basic models are models with input variables selected from the calendar (Table 6), weather (Table 2) and technical

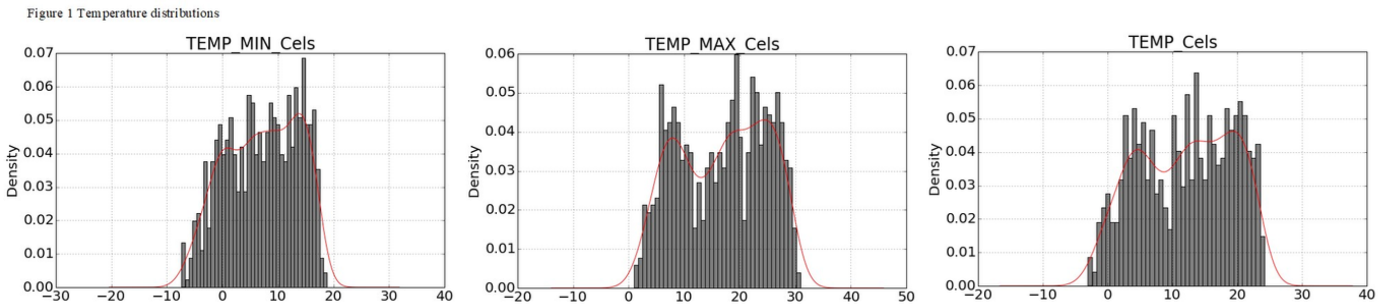

Figure 1: Temperature distribution

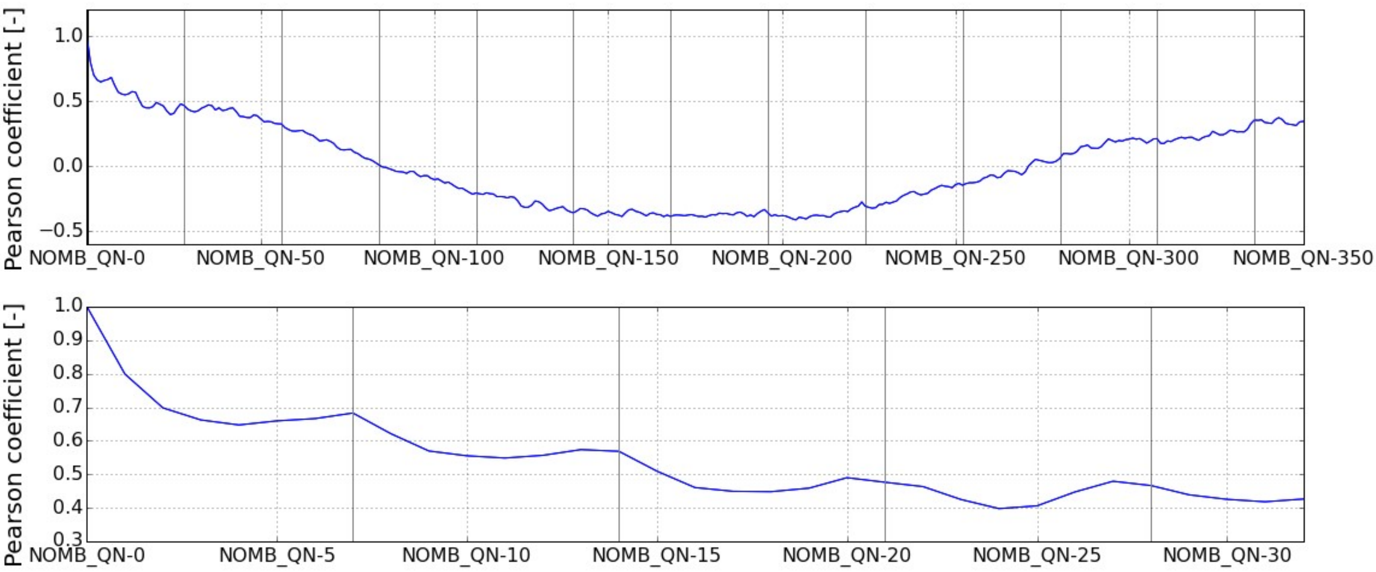

Figure 2: Pearson coefficient

data (Table 1). In advanced models gas market and macro-economic input variables were also considered. The distinction between basic and advanced models was made in light of the final prediction horizon, which is one year; hence model input variables have to be forecast for the same horizon. Basic model inputs (weather forecast, calendar and technical data) have lower forecasting uncertainty than advanced model inputs (macro-economic factors). Obviously, a one year weather forecast has forecast error, but the error is considerably smaller than the error in attempting to forecast gas prices, CO2 emission prices or other macro-economic factors. Therefore basic models are used to predict real behavior of UGS users in the future, while advanced models can be used to conduct WHAT-IF analyses. Advanced models predict UGS user behavior for various gas price ranges or macro-economic data. An attempt to develop advanced models was made for all users, but prediction error compared to basic models changed considerably for only two of them (user 1 and user 3).

Model validation was carried out using a 3-fold cross validation technique, where data was divided into folds according to years (for example, training fold: years 2013 and 2014, validation fol

2015)

The models developed were evaluated by three prediction accuracy coefficients:

$$R^2 = \frac{\sum_{t=1}^{n}(\widehat{y_t} - \overline{y})^2}{\sum_{t=1}^{n}(y_t - \overline{y})^2} \tag{6}$$

$$RMS = \sqrt{\frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \widehat{y_i})^2} \tag{7}$$

$$MAPE = \frac{\sum_{i=1}^{n}(| \frac{y_t - \widehat{y_t}}{y_t} |)}{n} * 100\% \tag{8}$$

where R^2 is the coefficient of determination and indicates the proportion of the variance in the dependent variable that is predictable from the independent variable. RMS is the square root of the arithmetic mean of the squares of the difference between the measured and predicted value, MAPE (mean absolute percentage error) is the measure of prediction accuracy, y_t is the measured value of output variable y in time t, $\widehat{y_t}$ is the predicted value of output y in time t and $\overline{y_t}$ is the mean value of y.

For model prediction performance and comparison between developed models, the RMS coefficient was selected as the most reliable, although the MAPE coefficient is more intuitive. This is because of the user nomination trends (variation above and below zero level) and for small nomination with bad prediction MAPE coefficient generates excessive values. The example can be seen in user 1 results where for model LB and LAA the RMS results are comparable (9.38 and 9.32) to R2 but the MAPE coefficient has significantly different values.

Table 8: USER1 modeling results

	MAPE	RMS	R^2
LB	76.4	150.5	0.92
LBA	75.4	156.6	0.92
LA	63.7	134.3	0.94
LAA	43.9	149.5	0.93
ANNB	48.5	130.5	0.94
ANNBA	52.5	144.6	0.93
ANNA	71.1	124.4	0.95
ANNAA	60.5	125.2	0.95

Table 9: User 1 model description

Input variables	TEMP_MAX, TEMP-7, LIMIT_ENTRY_MAX_QN
First layer	6
Second layer	6

3.1. USER 1

The results presented in Table 8 show that of the basic models the ANN-based model enjoys the best prediction accuracy. Moreover, the autoregression model gives a lower level of accuracy, which means the nomination actions are not repeatable. Advanced models for linear regression models gave improved prediction accuracy, but lower than for the ANNB model. The ANNA model enjoys the best prediction accuracy.

ANNB model architecture and input variables are presented in Table 9. Since the model is based on weather data, we can imply that the user uses this storage to compensate for varying energy and heat demand (temperature variables gives information on general energy demand) while the gas supply is constant. The LIMIT_ENTRY_MAX_QN input variable shows that the user submitting the gas flow nomination is constrained by limits signed in the contract.

The ANNA model has the same input variables as the ANNB model presented in Table 9 plus 'CO2_emission_price'. For the ANNA model RMS decreased to 6.1 compared to ANNB.

This case also shows that we can obtain a good prediction accuracy training model based on one season's data (summer, user nominate injection) and predict the other season's data (winter, user nominate withdrawal). Usually a neural network is not suitable for extrapolation, but by using the cross validation technique a network architecture has been chosen that enables capture of input-output dependencies with satisfactory results (R2 greater than 0.9). The measurements, predicted values and confidential intervals are presented in Fig. 3. Confidence intervals ($\alpha = 95\%$) are calculated in the range:

$$< y_{pred} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} : y_{pred} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} > \tag{9}$$

where: y_{pred} is predicted value, $t_{\frac{\alpha}{2}}$ is upper $\frac{\alpha}{2}$ critical value for t-distribution with n-1 degrees of freedom, s is residuals standard error and n is sample size.

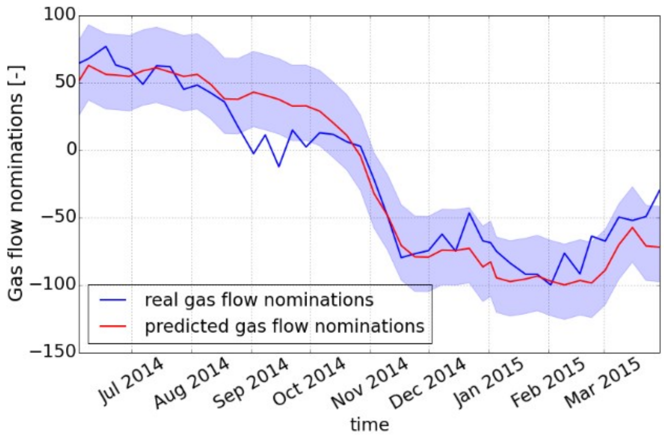

Figure 3: User 1 measurements and prediction comparison for ANNB model

Table 10: User 2 modeling results

	MAPE	RMS	R^2
LB	91.8	334	0.63
LBA	92.3	303	0.71
LA	81.6	351	0.61
LAA	81.4	329	0.66
ANNB	83.5	283	0.74
ANNBA	66.7	279	0.75
ANNA	84.5	298	0.72
ANNAA	75.0	308	0.7

3.2. USER 2

User results presented in Table 10 clearly show that basic models have better prediction accuracy than advanced models, which implies that UGS storage is used to compensate varying energy demand. Lack of prediction improvement after considering macro-economic and gas prices data shows that the user is not trading on the market.

ANN-based models have significantly better prediction accuracy than Linear Regression-based models. In ANN models additionally LIMIT_EXIT_MAX_QN, LIMIT_ENTRY_MAX_QN input variables were added, providing better artificial neural network performance in modeling non-linear dependencies. This also shows that the user is constrained by injection and withdrawal limits, which might mean that the user has signed a contract for insufficient limits to operate normally.

The best models were obtained for the ANNBA modeling technique, showing that user current nominations are based on the nomination made one week earlier. Model architecture and input variables are presented in Table 11. The measurements, predicted values and confidential intervals are presented in Fig. 4.

3.3. USER 3

The results presented in Table 12 show that the best prediction is obtained with advanced models, showing that by

Table 11: User 2 model description

Input variables	TEMP, TEMP_MIN, MAX_WIND_SPEED_m/s-7, LIMIT_EXIT_MAX_QN, LIMIT_ENTRY_MAX_QN, TEMP_MAX_Cels+3, WIND_SP_SQRT*TEMP+7, TEMP_MAX_Cels-5, TEMP_MAX-4, TEMP_MAX_Cels-1, WIND_SP_m/s_SQRT-1
First hidden layer size	16
Second hidden layer size	0

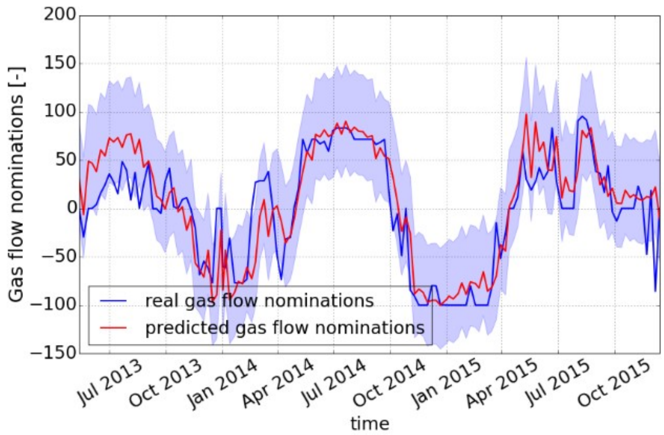

Figure 4: User 2 measurements and prediction comparison

adding macro-economic features we can significantly improve model performance. Basic models have lower accuracy than advanced models, implying user dependency on gas market and economic factors. For that reason the autoregressive modeling technique did not improve prediction accuracy (RMS is comparable).

The advanced model features are presented in Table 13. Input variables are related to: weather (temperature, wind speed), gas market (gas price), energy market (CO2 emission prices) and macro-economic (Oli_india_limited). Temperature input variables contained delayed data (from the user view it is historical weather data) and accelerated

Table 12: User 3, modeling results

	MAPE	RMS	R^2
LB	79.0	207.3	0.71
LBA	72.4	196.9	0.75
LA	49.5	166.4	0.82
LAA	49.1	167.8	0.83
ANNB	67.1	184.0	0.74
ANNBA	60.1	185.6	0.78
ANNA	67.9	220.0	0.69
ANNAA	73.5	219.2	0.69

Table 13: User 3 ANN model

Input variables	TEMP_MAX-3, Oli_IL, MAX_WIND_SPEED_m/s+3, Natural_Gas_Contract, MARCH, WIND_SP_SQRT*TEMP-9, TEMP_MIN_Cels+7, TEMP_MAX_Cels-4, CO2_high, WIND_SP_SQRT+3, TEMP_MAX_Cels-5, CO2_emission_price, Henry_Hub -4, WIND_SP_SQRT-15
First hidden layer size	6
Second hidden layer size	4

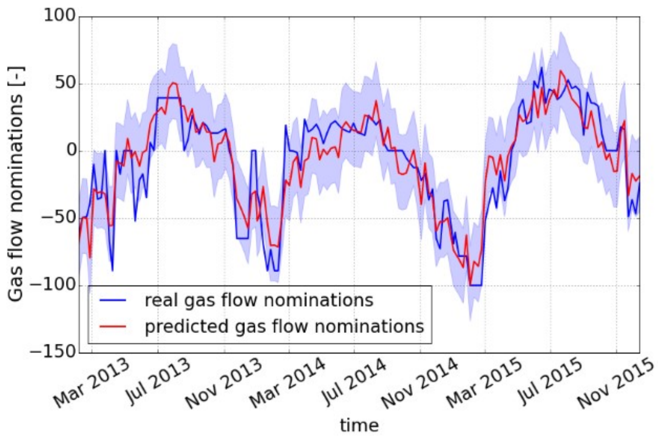

Figure 5: User 3 measurements and prediction comparison

data (from the user view this is the weather forecast), the user therefore relies on both current weather and forecasts. Therefore, the gas flow nomination prediction for this user with a one year horizon will require prediction of all input variables in the case of an advanced model.

The comparison between measurement and prediction with confidence interval is presented in Fig. 5.

3.4. User 4

For this user only basic models were developed. For advanced models, the algorithm did not choose any macro-economic or gas market features. The results are presented in Table 14. The RMS coefficient has similar values for all models, but the best seems to be the Artificial Neural Network.

Table 14: User 4 modeling results

	MAPE	RMS	R^2
LB	103.9	473.6	0.65
LBA	93.0	460.7	0.73
ANNB	84.1	442.5	0.74
ANNBA	84.5	447.7	0.74

Table 15: User 4 model description

Input variables	TEMP_MAX, TEMP-1, WIND_SPEED-30, LIMIT_EXIT_MAX_QN, LIMIT_ENTRY_MAX_QN, week_nb
First hidden layer size	6
Second hidden layer size	4

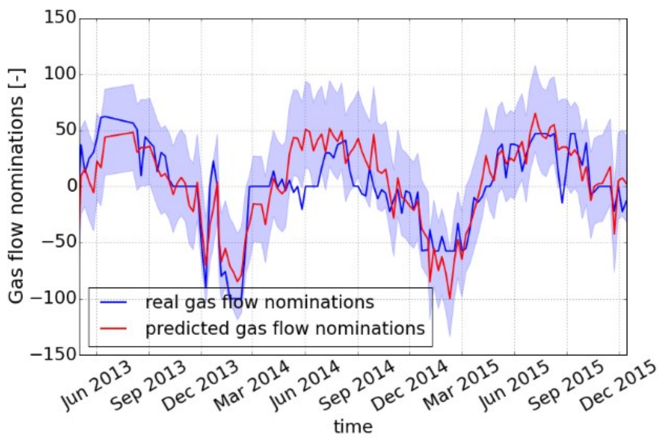

Figure 6: User 4 measurements and prediction comparison

The results presented in Table 15 imply that the user uses the gas storage to compensate irregular gas consumption in power plants due to the changing seasons (temperature input variables) and varying wind farm production (wind_speed input variable). In user model input variables a week_nb feature has been selected, implying that users' gas flow nominations are dependent on the current year period (user has similar nominations for the same months/weeks during many years).

The comparison between measurement and prediction with confidence area is presented in Fig. 6.

4. Conclusions

This paper presents an approach toward the gas flow nomination prediction of underground gas storage users. The UGS users were presented. Prediction models are based on weather, calendar, technical and macro-economic data. The results show that depending on the user profile, the choice of different model techniques provides the best outcomes. However, non-linear models usually have better prediction performance.

It has been shown that depending on user company profile and operation areas we can build prediction models on basic features (for user 2 and user 4 only basic models were developed) such as temperature variables (compensation of seasonal changes in gas demand) or wind variables (compensation of unstable electricity production in wind power plants). The temperature variable is considered as an input variable – with both delayed and accelerated values – showing that the user also relies on the weather forecast when defining gas flow nominations.

For users with basic models an additional input variable (representing constraints or individual independence) improves model performance, thus for user 2 LIMIT_EXIT_MAX_QN, LIMIT_ENTRY_MAX_QN variables improved the model, revealing that the user is constrained by factors signed in the contract and for user 4 the week_nb input variable showed user dependency on the current year season. For user 1 the CO2 emission process factor improved the model (based on temperature, wind and limits inputs), which might represent user operation area, i.e., energy and heat supply. A great distinction between basic and advanced models is apparent with user 3, where macro-economic data like gas process, CO2 emission process, historical prices for Oil India Limited all improved model accuracy. User 3 is the only user among those analyzed where the Linear regression based model enjoys better accuracy than the Artificial Neural Network-based models. All user auto regression did not make a significant improvement. The Artificial Neural Network enjoys better performance due to users' dependency on non-linear variables, such as LIMIT_EXIT_MAX_QN or week_nb.

The models developed are intended to be used as support for facility operation decisions and for optimization of product portfolios.

References

[1] O. Flanigan, Underground gas storage facilities: Design and implementation, Elsevier, 1995.

[2] G. Ding, C. Li, J. Wang, H. Xu, Y. Zheng, Q. Wanyan, Y. Zhao, The status quo and technical development direction of underground gas storages in china, Natural Gas Industry B 2 (6) (2015) 535–541.

[3] C. Budny, R. Madlener, C. Hilgers, Economic feasibility of pipe storage and underground reservoir storage options for power-to-gas load balancing, Energy Conversion and Management 102 (2015) 258–266.

[4] E. N. Escobar, G. R. Arteaga Mora, A. G. Kemp, et al., Underground natural gas storage in the uk: Business feasibility. case study, in: SPE EUROPEC/EAGE Annual Conference and Exhibition, Society of Petroleum Engineers, 2011.

[5] B. Žlender, S. Kravanja, Cost optimization of the underground gas storage, Engineering Structures 33 (9) (2011) 2554–2562.

[6] R. Danel, L. Otte, V. Vancura, M. Řepka, Monitoring and balance of gas flow in underground gas storage, Procedia earth and planetary science 6 (2013) 485–491.

[7] K. Wojdan, B. Ruszczycki, D. Michalk, K. Swirski, Method for simulation and optimization of underground gas storage performance, Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles 69 (7) (2014) 1237–1249.

[8] A. S. Bagci, E. Ozturk, Performance prediction of underground gas storage in salt caverns, Energy Sources, Part B 2 (2) (2007) 155–165.

[9] T. Ligen, W. Jieming, D. Guosheng, S. Shasha, Z. Kai, S. Junchang, G. Kai, B. Fengjuan, Downhole inflow-performance forecast for underground gas storage based on gas reservoir development data, Petroleum Exploration and Development 43 (1) (2016) 138–142.

[10] H. Aras, N. Aras, Forecasting residential natural gas demand, Energy Sources 26 (5) (2004) 463–472.

[11] F. Gorucu, Artificial neural network modeling for forecasting gas consumption, Energy Sources 26 (3) (2004) 299–307.

[12] F. Gümrah, D. Katircioglu, Y. Aykan, S. Okumus, N. Kilincer, Modeling of gas demand using degree-day concept: case study for ankara, Energy sources 23 (2) (2001) 101–114.

[13] E. F. Sánchez-Úbeda, A. Berzosa, Modeling and forecasting industrial end-use natural gas consumption, Energy Economics 29 (4) (2007) 710–742.

[14] R. Gutiérrez, A. Nafidi, R. G. Sánchez, Forecasting total natural-gas consumption in spain by using the stochastic gompertz innovation diffusion model, Applied Energy 80 (2) (2005) 115–124.

[15] Z. Wadud, H. S. Dey, M. A. Kabir, S. I. Khan, Modeling and forecasting natural gas demand in bangladesh, Energy Policy 39 (11) (2011) 7372–7380.

[16] L.-M. Liu, M.-W. Lin, Forecasting residential consumption of natural gas using monthly and quarterly time series, International Journal of Forecasting 7 (1) (1991) 3–16.

[17] National Centers for Environmental Information (NCEI) formerly known as National Climatic Data Center (NCDC) | NCEI offers access to the most significant archives of oceanic, atmospheric, geophysical and coastal data., https://www.ncdc.noaa.gov/.
URL `https://www.ncdc.noaa.gov/`