Applied Energy 228 (2018) 1091–1110

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Forecasting day-ahead high-resolution natural-gas demand and supply in Germany

Ying Chen[a], Wee Song Chua[a,*], Thorsten Koch[b,c]

[a] Department of Statistics & Applied Probability, Block S16, Level 7, 6 Science Drive 2, Faculty of Science, National University of Singapore, Singapore 117546, Singapore
[b] Department of Mathematics, Technische Universität Berlin, Str. des 17. Juni 136, 10623 Berlin, Germany
[c] Mathematical Optimization Department at Zuse Institute Berlin, Takustr. 7, 14195 Berlin, Germany

HIGHLIGHTS

- Develop a new functional model to forecast day-ahead natural gas-flow in Germany.
- Incorporate functional exogenous variables in the modeling.
- Strong evidence that the proposed model outperforms several alternative models.
- Improve forecast accuracy by up to 5-fold.
- Results show that nominations and lagged 1-day flow are important predictors.

ARTICLE INFO

Keywords:
Natural gas
Forecasting
Functional time-series
Autoregressive
Demand and supply

ABSTRACT

Forecasting natural gas demand and supply is essential for an efficient operation of the German gas distribution system and a basis for the operational decisions of the transmission system operators. The German gas market is moving towards more short-term planning, in particular, day-ahead contracts. This increases the difficulty that the operators in the dispatching centre are facing, as well as the necessity of accurate forecasts. This paper presents a novel predictive model that provides day-ahead forecasts of the high resolution gas flow by developing a Functional AutoRegressive model with eXogenous variables (FARX). The predictive model allows the dynamic patterns of hourly gas flows to be described in a wide range of historical profiles, while also taking the relevant determinants data into account. By taking into account a richer set of information, FARX provides stronger performance in real data analysis, with both accuracy and high computational efficiency. Compared to several alternative models in out-of-sample forecasts, the proposed model can improve forecast accuracy by at least 12% and up to 5-fold for one node, 3% to 2-fold and 2-fold to 4-fold for the other two nodes. The results show that lagged 1-day gas flow and nominations are important predictors, and with their presence in the forecast model, temperature becomes insignificant for short-term predictions.

1. Introduction

Natural gas is a key energy resource for Germany. In 2012, about 23% of the energy supply of Germany was provided by natural gas (see [1]). The German high-pressure transport pipeline network has a length of about 35,000 km. This network is used to transport the gas either through the country or to the local distribution networks of more than 300,000 km length. Given that gas on average travels rather slowly, e.g., 25 km/h, even in high-pressure pipelines, it becomes evident that an efficient high-precision, high-frequency forecasting of local supplies and demands is essential for efficient operation of the network. The German gas market is moving towards more short-term planning, in particular, day-ahead contracts. This increases the difficulty that the operators in the dispatching centre are facing. We propose a day-ahead predictive model, namely the Functional AutoRegressive model with eXogenous variables (FARX) to study the dynamic structure of the high-resolution natural gas flows in a unified framework. Our model captures both the data's property of serial cross-dependence and causal relationship with meteorological and economic data.

Although natural gas forecasting has become a fundamental input to network operators' scheduling and decision-making mechanisms, there has been little research reported in the literature, at least not to the same extent as Electricity Price Forecasting (EPF). Soldo [2] presents a survey of forecasting natural gas consumption, where the overall gas

* Corresponding author.
E-mail address: a0054070@u.nus.edu (W.S. Chua).

https://doi.org/10.1016/j.apenergy.2018.06.137
Received 13 April 2018; Received in revised form 20 June 2018; Accepted 29 June 2018

flow is dependent on meteorological and economic factors such as weather, working routines and regulations. Chen et al. [3] use geological parameters to estimate the adsorbed shale gas content via statistical learning. Crow et al. [4] propose a bottom-up model of natural gas supply by simulating investment and operating decisions in the upstream gas industry triggered in response to investors expectations of future gas prices. Qiao et al. [5] build a comprehensive system model of a coupled natural gas and electricity network and conclude that the uncertainty of wind power has a significant impact on the steady-state operation of natural gas systems.

In view of natural gas being widely used for heating purposes, it is only natural to consider temperature as a key indicator in forecasting gas demands. Stoll and Wiebauer [6] show that there is a significant relationship between historical natural gas spot prices and temperature. Berndt and Watkins [7] investigate the effect of temperature variation on natural gas demand. On top of the direct impact on the demand, diurnal patterns and seasonal variations of natural gas are caused by temperature changes, in addition to the nature of business intensity and economic routine. Vitullo et al. [8] forecast natural gas demand using temperature, wind, previous day demand, day of the week, and holidays as inputs in their model. Similar causality with the previous inputs is found in other energy forecast such as electricity (e.g. [9]), and wind (e.g. [10]). Moreover, market regulation plays a role in gas demand and supply in Germany, where suppliers and consumers of natural gas have to nominate the expected gas purchase quantities for certain nodes one day in advance. Nominations can be amended and cancelled by the customers until three hours before the gas is transferred, yet nominations have not been well studied due to limited data availability. Econometric models are eventually driven by utilising this causal dependence, forecasting gas flows with exogenous variables. For example, Berndt and Watkins [7] forecast the demand for natural gas with temperature as exogenous. Khan [11] finds that per capita real income exerts a positive and significant impact on the natural gas consumption, while Gorucu [12] forecasts gas consumption using exogenous variables such as temperature, selling price of gas, number of customers, and exchange rate.

Like the electricity price, gas flow has a high time resolution and exhibits a stochastic behaviour over time, indicating the possibility of adapting the popular EPF methods and models for gas forecasting. In the EPF literature, the AutoRegressive (AR) model is popular given its generally good performance (see [13]). The AR model is often implemented separately for each hour with or without exogenous determinant, leading to 24 forecasting specifications. Nogales et al. [14] forecast next day electricity prices using past observations and include demand as an exogenous variable, while Contreras et al. [15] use a more general AR model, namely Autoregressive Integrated Moving Average (ARIMA) models, to predict the electricity price on the next day. In gas forecasting, Potocnik et al. [16] and Cardoso and Cruz [17] use ARX and ARIMA respectively to forecast natural gas consumption, while Sharmin and Khan [18] and Al-Fattah [19] use ARIMA to forecast natural gas production. It is worth noting that although the multi-model forecasting generally delivers good forecast accuracy in real-world data analysis, it ignores the rich information available in the multiple series, namely the lead-lag dependence among the multiple hourly series. Vector AR (VAR) and Vector ARX (VARX) models can describe the cross-dependence, given the 24 hourly resolution. However, there are hundreds of parameters and an overfitted model might describe random noise instead of the underlying relationship, thus leading to poor forecast accuracy.

Recent developments in functional data analysis pave the way for jointly analysing the high-dimensional dependent stochastic processes efficiently. On each day, the multiple hourly gas flows can be smoothed as a continuous curve that naturally inherits the serial cross-dependence in the original hourly flows. Instead of 24 series, one for each hour, there is only one single series of daily curves to be analysed in the Functional AutoRegressive (FAR) framework. Although the curves are defined in an infinite parameter space, the estimation of FAR is conducted in a projected limited parameter space reflecting the essential dependence, which avoids the overfitting problem as mentioned above. In particular, Mourid and Bensmain [20] derive a maximum likelihood (ML) estimator of the FAR model on the sieve approximation, to avoid the inverse problem in functional Yule-Walker (YW) estimation proposed by Bosq [21] or other competing YW estimators evaluated by Kokoszka and Zhang [22]. Moreover, Chen and Li [23] extend it and derive a closed-form ML solution for the adaptive FAR model that can be used in both stationary and nonstationary situations.

Our main contribution lies in the development of a day-ahead predictive model for high resolution gas data by utilizing comprehensive information. The flexibility of the FARX model allows describing the dynamic patterns of hourly gas flows in a wide range of historical profiles while simultaneously taking the relevant determinants data into account. The implementation is computationally efficient with a closed-form estimation, for which we also provide a proof of consistency. This work is part of a joint project with Germanys biggest transmission system operator (TSO), where we are building an advanced prediction system, providing day-ahead hourly forecasts for various nodes with different features in the gas pipeline network. FARX is the forecast method considered to be used in the forward-looking distribution system. In real applications, we perform an extensive empirical forecasting experiment on hourly resolution flows of three large nodes in the gas transmission network. We find that the FARX model significantly outperforms most of the alternative models in the out-of-sample day-ahead forecast experiments. The FARX model improves forecast accuracy by at least 12% and up to 5-fold for one node, 3% to 2-fold and 2-fold to 4-fold for the other two nodes. We also find that the contribution of temperature is moderate for this particular case. It shows that the FARX model, taking into account richer information, provides stronger performance in real data analysis, with both accuracy and high computational efficiency. As such, our study also contributes to the literature by comparing the performance of the FARX model with a number of popular predictive models in univariate, multivariate, and functional domains with and without exogenous variables and seasonality.

It is worth noting that recent EPF literature has seen a rise in adoption of the artificial neural network (ANN) approach in modeling. Bento et al. [24] use ANN and a wavelet transform approach for short-term price forecasting, while Panapakidis and Dagoumas [25] and Keles et al. [26] adopt ANN combined with a clustering algorithm for day-ahead electricity price forecasts. Lago et al. [27] propose a deep learning framework to forecast electricity prices and compared this with 27 common approaches. Likewise in the gas literature, there are multiple papers on using ANN models to forecast natural gas demands or consumptions (e.g. [17]), as well as hybrid models such as a combination of ANN forecasters (e.g. [28]), a combination of the commonly used multiple linear regression and ANN (e.g. [8,29]), and a combination of time series methods and ANN (e.g. [30]). Though not shown in the paper, we also adopt an ANN approach in the forecast of gas flow, but the results are not satisfactory when compare to the models considered in this paper. Furthermore, the focus of this paper is on the proposed FARX model and its performance compare to the alternative autoregressive models, hence, the results from the ANN approach are omitted.

The paper is structured as follows. Section 2 presents the FARX model, its closed-form estimator, and the alternative models. In Section 3, we describe the gas flow data, report the analytical results of the natural gas flows in Germany and provide discussions. Section 4 gives