

MILENA PETKOVIC<sup>1</sup>, NAZGUL ZAKIYEVA<sup>2</sup>, JANINA  
ZITTEL<sup>3</sup>

## **Statistical analysis and modeling for detecting regime changes in gas nomination time series**

---

<sup>1</sup>  0000-0003-1632-4846  
<sup>2</sup>  0000-0001-9106-9916  
<sup>3</sup>  0000-0002-0731-0314

Zuse Institute Berlin  
Takustr. 7  
14195 Berlin  
Germany

Telephone: +49 30 84185-0  
Telefax: +49 30 84185-125

E-mail: [bibliothek@zib.de](mailto:bibliothek@zib.de)  
URL: <http://www.zib.de>

ZIB-Report (Print) ISSN 1438-0064  
ZIB-Report (Internet) ISSN 2192-7782

# Statistical analysis and modeling for detecting regime changes in gas nomination time series

Milena Petkovic, Nazgul Zakiyeva, and Janina Zittel

Zuse Institute Berlin, Applied Algorithmic Intelligence Methods Department,  
Takustraße 7, 14195, Berlin  
petkovic@zib.de   zakiyeva@zib.de   zittel@zib.de

**Abstract.** As a result of the legislation for gas markets introduced by the European Union in 2005, separate independent companies have to conduct the transport and trading of natural gas. The current gas market of Germany, which has a market value of more than 54 billion USD, consists of Transmission System Operators (TSO), network users, and traders. Traders can nominate a certain amount of gas anytime and anywhere in the network. Such unrestricted access for the traders, on the other hand, increase the uncertainty in the gas supply management. Some customers' behaviors may cause abrupt structural changes in gas flow time series. In particular, it is a challenging task for the TSO operators to predict gas nominations 6 to 10 hours-ahead. In our study, we aim to investigate the regime changes in time series of nominations to predict the 6 to 10 hours-ahead of gas nominations.

**Keywords:** regime switching, nonlinear time series, gas nomination forecast

## 1 Introduction

Natural gas plays as key energy source in Europe. Due to the legislation for gas markets introduced by the European Union in 2005, gas traders can independently nominate a certain amount of gas in any entry or exit point in the network. The amount of gas, either input or off-take, can be recorded or changed by the trader anytime up to 3am on each day. Such unrestricted access for the traders increases the uncertainty in the gas network management [2]. In particular, abrupt pattern changes of some customers' behaviors make it challenging for the gas network operators to predict 6 to 10 hours-ahead gas nominations.

From the point of view of our industry partner, one of Germany's biggest TSOs, Open Grid Europe [3], the objective is to accurately predict the final gas nominations at 6am to 10am starting at midnight. It is the most challenging period of prediction horizon in gas nomination forecast, as the final gas nomination for 6am can be abruptly change around 2am or 3am. Autoregression based time series models are popular in energy prediction due to its interpretability and accurate forecasting compared to black-box machine learning methods [1, 7]. In our

study we apply Markov Switching Autoregression model with Exogenous (MS-ARX) to study the regime changes in customer’s behaviour pattern and predict 6 to 10 hours ahead final gas nominations. MS-ARX detects 3 regimes and shows superior multi-step ahead prediction performance compared to the benchmark and linear models. 93% of days in the out-of-sample period, MS-ARX provides effective predictions for 6am to 10am predictions.

## 2 Data

We study the regime changes in gas nomination time series in the German gas pipeline network. We consider the dataset of one customer nominating gas amount in seven different entry points, which creates the most complex dynamics for multi-step ahead prediction among others. The dataset covers 678 days (22 months and 6 days) from March Y1 to January Y2 with 16,272 observations, where Y1 and Y3 denotes the years of the dataset.

Figure 1a displays the aggregated final gas nomination time series of one customer at 7 entry points. The dynamic pattern of the time series shows sudden changes within the the first 6 to 10 hours from midnight. The short period of sudden increases and drops in time series may challenge the modeling and multi-step-ahead (6 to 10 hours-ahead) predictions of final gas nomination. It motivates us to consider regime changes in the time series modeling, as the customer’s behaviour in nominating gas may be dependent on the unobservable discrete state variables.

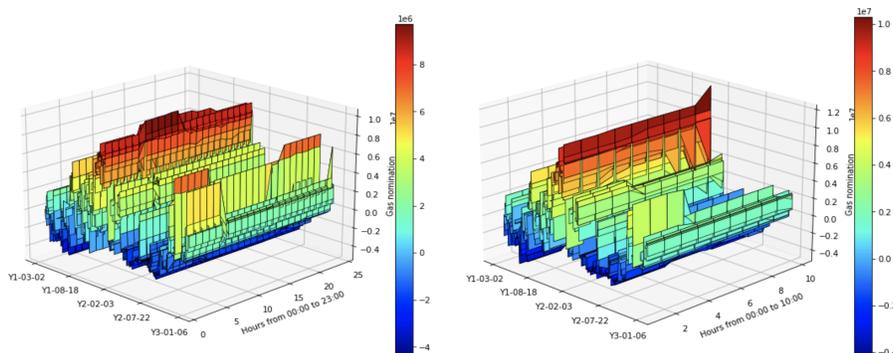
Gas nomination history is an additional time series dataset which contains nomination history for every final hour in a day. Figure 1b illustrates the gas nomination history up to 10 hours of offset for daily final gas nominations for 6am from March Y1 and January Y3. The graph shows the amount of the gas nominated for 6am 1 to 10 hours in advance every day. It also has abrupt time series changes around three hours before 6am. Figure 2 shows the cross-dependence between the final nominations at 24 hours and their history up to 10 hours of offset. Although the cross-dependence between hourly final nominations and their history show high correlation, 6am final nomination and its history start decreasing after 4 hours of offset, which is the nomination set at 2am. And similar pattern is observed for 7am to 12pm.

## 3 Method

Let  $y_t, t = 1, \dots, T$  denote a time series of T univariate observations. Observable stochastic process  $y_t$  is

$$y_t = \mu_{S_t} + \phi_{S_t}x + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_{S_t}^2), \quad (1)$$

where  $x$  is a vector observed explanatory variables and  $\phi_{S_t}$  is a vector unknown parameters depending on hidden discrete stochastic process  $S_t$ , [4, 5]. The probability of moving from regime  $i$  to regime  $j$ :  $P(S_t = j | S_{t-1} = i) = p_{ij}$ ,  $i, j =$



(a) Final gas nomination time series. (b) Gas nomination history for 6am.

Fig. 1: Final gas nomination time series and gas nomination history between March 2 Y1 and January 6 Y3, where Y1 and Y3 denotes the years of the dataset.

$0, 1, \dots, K-1$  called transition probabilities, with  $p_{i,0} + p_{i,1} + \dots + p_{i,K-1} = 1$ ,  $i = 0, 1, \dots, K-1$ . In this process the current regime depends only on the regime before and it is called a Markov process. With given past observations  $F_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$ , we estimate the transition probabilities  $p_{ij}$  by maximizing the likelihood function with respect to the unknown parameters  $\theta = (p_{i,j}, \phi_{S_t}, \mu_{S_t}, \sigma_{S_t}^2)$ ,  $i, j = 0, \dots, K-1$  and the transition probabilities  $p_{i,j}$ . Taking the log of the likelihood function  $l(\theta) = f(\theta|F_{t-1})$  and using the chain rule for the conditional probability, we have the log-likelihood function as follows,  $l(\theta) = \sum_{t=1}^T l_t(\theta) = \sum_{t=1}^T (\ln \sum_{S_t=0}^{m_s} \sum_{S_{t-1}=0}^{m_s} f(y_t, S_t, S_{t-1}|F_{t-1})) = \sum_{t=1}^T (\ln \sum_{S_t=0}^{m_s} \sum_{S_{t-1}=0}^{m_s} f(y_t|S_t, S_{t-1}, F_{t-1})P(S_t, S_{t-1}|F_{t-1}))$  The conditional probability density function for the observations  $y_t$  given the state variables  $S_t, S_{t-1}$  and the previous observations

$$f(y_t|S_t, S_{t-1}, F_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_{S_t}^2}} \exp \left\{ -\frac{[y_t - \mu_{S_t} - \phi_{S_t}x]^2}{2\sigma_{S_t}^2} \right\},$$

as  $\epsilon_t = y_t - \mu_{S_t} - \phi_{S_t}x \sim N(0, \sigma_{S_t}^2)$ . Using the chain rule for conditional probabilities and the Markov property  $P(S_t|S_{t-1}, F_{t-1}) = P(S_t|S_{t-1})$ , we rewrite the conditional joint probability  $P(S_t, S_{t-1}|F_{t-1})$  as

$$P(S_t, S_{t-1}|F_{t-1}) = P(S_t|S_{t-1}, F_{t-1})P(S_{t-1}|F_{t-1}) = P(S_t|S_{t-1})P(S_{t-1}|F_{t-1}).$$

The time dependent state probabilities  $P(S_{t-1}|F_{t-1})$  are updated by Kim algorithm [6]. Finally, with the estimated parameters, the  $h$  step-ahead prediction is calculated as the weighted average of the predictions in each regime,  $\hat{y}_{t+h} = \sum_{j=0}^{K-1} (\mu_{S_{t+h}} + \phi_{S_{t+h}}x)P(S_{t+h} = j|F_t)$ .  $P(S_{t+h} = j|\Phi_t)$  is a smoothed probability that observation at time  $t+h$  is in state  $j$  given the history  $F_t = \{y_t, \dots, y_1\}$ .

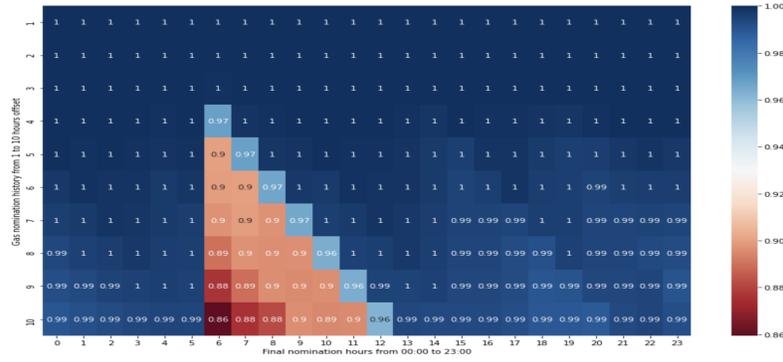


Fig. 2: Correlation between hourly final gas nomination and the gas nomination history offset hours.

## 4 Forecast setup and evaluation

We predict final gas nomination time series described in Section 2. We make 6 to 10 hours-ahead out-of-sample forecast in real time starting from 00:00 on January 1 Y2 to January 6 Y3, a total of 373 days. In the training-validation sample we choose the optimal hyperparameters for the training the model with sliding window method. Then we estimate the model weights at each point in the out-of-sample period.

For an  $h$ -step ahead forecast, where  $h = 6, 7, 8, 9, 10$ , we apply the MS-ARX model with 3 regimes and 3 variables including the  $y_{t_0}, y_{t_0-24+h}$  and  $x_{t_0}$ , denoting the last observed gas nomination, 24-h hours before gas nomination and history of gas nomination at time  $t_0$ , respectively, where  $t_0$  is the last observed point. Updating the model using the whole observed train sample, we move forward predicting the rest of the test sample one period at a time.

As alternative models, we apply three benchmark methods namely, Base ( $y_{t_0}$ ), Base ( $y_{t_0-24+h}$ ), and Reference ( $x_{t_0}$ ). Furthermore, we consider Autoregression model with the exogenous (ARX) using the three variables. We train the ARX model using the last 7 days of rolling window, as it gave the optimal out-of-sample prediction results in the training-validation period.

We evaluate the relative forecast accuracy according to average root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) for the  $h$ -step-ahead predictions. The smaller the average error, the better accuracy is obtained by the forecast model. We obtain hourly  $h$ -step-ahead predicted gas flow series  $\hat{Y}_{t+h}$ . The  $h$ -step-ahead prediction performance is evaluated over the forecasting period. The RMSE, MAE, and MAPE are obtained as  $\text{RMSE} = [\sum_{t \in T_2} (\hat{y}_{t+h} - y_{t+h})^2 / |T_2|]^{1/2}$ ,  $\text{MAE} = \sum_{t \in T_2} |\hat{y}_{t+h} - y_{t+h}| / |T_2|$ , and  $\text{MAPE} = \sum_{t \in T_2} |(\hat{y}_{t+h} - y_{t+h}) / y_{t+h}| / |T_2|$ , for  $h = 6, \dots, 10$ , where  $|T_2|$  is the length of forecasting period  $T_2$ .

Furthermore, we apply traffic light evaluation criteria. As discussed in Section 2, the extreme sudden changes in the final gas nomination time series occur between 2am to 5am. It complicates predicting the 6am to 10 am final gas nominations using the observations only until 12am. The proportion of green days are evaluated with  $MAE < 10^6 W$  and  $MAPE < 0.1$ , yellow days with  $MAE < 2 * 10^6 W$  or  $MAPE < 0.2$ , and red days with  $MAE > 10^6 W$  and  $MAPE > 0.1$  in predicting 6am to 10am final gas nominations trained with the time series until 12am.

## 5 Forecast results

We demonstrate the multi-step ahead forecasting results of final gas nomination of a difficult customer in German gas transmission network. Table 1 shows the average forecast errors, denoted as aRMSE, aMAE and aMAPE, by the MS-ARX and the alternative models for the 6 to 10 hours-ahead gas nomination prediction. Table 1 compares the out-of-sample predictions done at 12am for 6am to 10am using the traffic light criteria. Overall, the results demonstrate the advantages of the regime changes assumption in improving the multi-step ahead final gas nomination prediction. The MS-ARX performs as the most accurate forecast model with the smallest average forecast errors for multi-step ahead predictions. It illustrates that the assumption of the regime changes in the final gas nomination improves the forecast errors of the linear ARX model from aRMSE 5.18, MAE 2.35 and MAPE 1.65 to RMSE 3.85, MAE 1.48 and MAPE 0.67 for 6 to 10 hours-ahead gas nomination forecast. Compared to the Reference ( $x_{t_0}$ ), which is the best performing model within the benchmark models, the MS-ARX improves its prediction errors from aRMSE 5.88, MAE 2.03 for 6 to 10 hours-ahead predictions. Table 1 shows the MS-ARX model outperforms all the alternative models with the maximum percentage of green days (93.01%) and minimum percentage of the red days (2.42%). It is worth mentioning that the MS-ARX shows not only the accurate prediction results on overall h-step-ahead forecasts, but its significant advantage is shown in predicting the final nomination in the most difficult hours (6am to 10am). In summary, it is useful to consider the regime changes assumption for accurate and stable multi-step ahead final gas nominations forecasts.

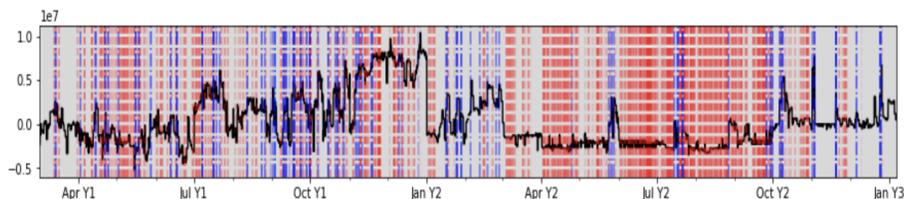


Fig. 3: Aggregated hourly final gas nomination in 3 estimated regimes, regimes 1, 2 and 3 are shown in grey, blue, red, respectively.

	MS-ARX	ARX	Base ( $y_{t_0}$ )	Base ( $y_{t_0-24+h}$ )	Reference ( $x_{t_0}$ )
aRMSE ( $10^6$ W)	<b>3.85</b>	5.18	7.66	11.86	5.88
aMAE ( $10^6$ W)	<b>1.48</b>	2.35	2.82	6.13	2.03
aMAPE	0.67	1.65	1.43	4.66	<b>0.55</b>
Green	<b>93.01%</b>	87.37%	82.26%	81.99%	72.31%
Yellow	<b>4.57%</b>	7.80%	10.48%	10.75%	20.70%
Red	<b>2.42%</b>	4.84%	7.26%	7.26%	6.99%

Table 1: Average forecast errors for h-step ahead forecast,  $h = 6, 7, 8, 9, 10$ , and traffic light criteria for 6am to 10am nomination forecast in particular. Best forecast result is marked in bold-face.

## 6 Conclusion

We apply the MS-ARX model to predict 6 to 10 hours-ahead final gas nominations in the German high-pressure gas pipeline network. It demonstrates the 3 regime periods in the hourly gas nomination time series and shows a superior out-of-sample forecast performance over linear and benchmark prediction models.

## 7 Acknowledgement

The work for this article has been conducted within the Research Campus Modal funded by the German Federal Ministry of Education and Research (fund numbers 05M14ZAM, 05M20ZBM)

## References

- [1] Marek Brabec et al. “A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers”. In: *International Journal of Forecasting* 24.4 (2008), pp. 659–678.
- [2] A. Fügenschuh et al. “Mathematical optimization for challenging network planning problems in unbundled liberalized gas markets”. In: *Energy Systems* 5 (2014), pp. 449–473.
- [3] Open Grid Europe GmbH. <https://www.oge.net/>.
- [4] James D Hamilton. “A new approach to the economic analysis of nonstationary time series and the business cycle”. In: *Econometrica: Journal of the econometric society* (1989), pp. 357–384.
- [5] James D Hamilton. “Regime switching models”. In: *Macroeconometrics and time series analysis*. Springer, 2010, pp. 202–209.
- [6] Chang-Jin Kim. “Dynamic linear models with Markov-switching”. In: *Journal of Econometrics* 60.1-2 (1994), pp. 1–22.
- [7] Teresa Kurek, Konrad Wojdan, and Konrad Świrski. “Long-term prediction of underground gas storage user gas flow nominations”. In: *Journal of Power Technologies* 99.4 (2020), pp. 272–280.