



NAZGUL ZAKIYEVA<sup>1</sup>, MILENA PETKOVIC<sup>2</sup>

# High-dimensional high-frequency time series prediction with a mixed integer optimisation method

---

<sup>1</sup>  0000-0001-9106-9916

<sup>2</sup>  0000-0003-1632-4846

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

# High-dimensional high-frequency time series prediction with a mixed integer optimisation method

Nazgul Zakiyeva<sup>1</sup> and Milena Petkovic<sup>1</sup>

<sup>1</sup> Technische Universität Berlin, Chair of Software and Algorithms for Discrete Optimization, Straße des 17. Juni 135, 10623 Berlin, Germany

`zakiyeva@zib.de`

<sup>2</sup> Zuse Institute Berlin, Applied Algorithmic Intelligence Methods Department, Takustraße 7, 14195 Berlin, Germany

`petkovic@zib.de`

**Abstract.** We study a functional autoregressive model for high-frequency time series. We approach the estimation of the proposed model using a Mixed Integer Optimisation method. The proposed model captures serial dependence in the functional time series by including high-dimensional curves. We illustrate our methodology on large-scale natural gas network data. Our model provides more accurate day-ahead hourly out-of-sample forecast of the gas in and out-flows compared to alternative prediction models.

**Keywords:** Functional autoregression, forecasting, mixed-integer programming.

## 1 Introduction

In the era of Big Data, large-scale time series data are observed in high dimensions and frequency. Functional Autoregression based models have been popular in modeling high-frequency time series [3,2,5]. Moreover, with the advancement of mathematical programming methods, it has been possible to solve large-scale real-life problems to optimality in a few seconds [1,4]. We propose to estimate the Functional Autoregression model with a Mixed Integer Optimization method and demonstrate its application in predicting gas flows at more than 1000 gas nodes in the German gas transmission network. For efficient energy transmission, it has been an essential task for gas transmission operators to accurately forecast gas flows in the network for the next 24 hours [4,5]. The results show that our model outperforms benchmark time series models for most hours and nodes in the network satisfying the optimal solution.

## 2 Method

### 2.1 Functional Autoregression

Let  $T$  be the number of time period. We denote  $\{Y_t(\tau)\}_{t=1}^T$  as a series of  $T$  random curves taking values in the Hilbert space  $\mathcal{H}$  on the continuous domain  $\tau \in [0, 1]$ . The  $p$  lagged curves are denoted as  $\{Y_{t-1}(\tau), \dots, Y_{t-p}(\tau)\}$ .

We present the functional autoregression (FAR) model, proposed in [3]. In particular, the FAR(1) is defined as:

$$Y_t(\tau) = \mu(\tau) + \int_0^1 \beta(\tau - s)Y_{t-1}(s)ds + \epsilon_t(\tau), \quad (1)$$

where the serial dependence of the functional response on the lagged values of its own series are measured by  $\beta(\cdot)$  which are square-integral regression parameter functions in  $\mathcal{H}$ . The innovation function  $\epsilon_t(\tau)$  is strong  $\mathcal{H}$ - white noise with zero mean and finite second moment  $E\|\epsilon_t(\tau)\|^2 < \infty$ . The norm  $\|\cdot\|$  is induced by the inner product  $\langle \cdot, \cdot \rangle$  of  $H$ . We further assume that the serial dependence is controlled by  $\beta(\cdot)$ . The autoregressive (AR) term controls the serial by the bounded linear AR operators  $\mathcal{H}$  to  $\mathcal{H}$ , with the AR operators specified as the regression functions  $\beta(\cdot)$ .

We represent the functional terms, i.e. the functional variables and the functional parameters, in the basis of  $L^2([0, 1])$  given by the trigonometric functions:

$$\Phi_0 = I_{[0,1]}, \Phi_{2k}(\tau) = \sqrt{2} \cos(2\pi k\tau), \Phi_{2k-1}(\tau) = \sqrt{2} \sin(2\pi k\tau)$$

for  $k \in \mathbb{Z} \setminus \{0\}$ , which have the orthogonal properties, see details in [2]. Using the properties, we derive the relationship equations between the Fourier coefficients as follows:

$$\begin{aligned} Y_t(\tau) &= a_{t,0} + \sum_{k=1}^{\infty} [b_{t,k}\Phi_{2k-1}(\tau) + a_{t,k}\Phi_{2k}(\tau)], \\ \mu(\tau) &= p_0 + \sum_{k=1}^{\infty} [p_k\Phi_{2k-1}(\tau) + q_k\Phi_{2k}(\tau)], \\ \beta(\tau) &= u_0 + \sum_{k=1}^{\infty} [v_k\Phi_{2k-1}(\tau) + u_k\Phi_{2k}(\tau)], \\ \epsilon_t(\tau) &= \eta_{t,0} + \sum_{k=1}^{\infty} [\omega_{t,k}\Phi_{2k-1}(\tau) + \eta_{t,k}\Phi_{2k}(\tau)], \end{aligned}$$

where  $a_{t,0}, b_{t,k}, a_{t,k}$  are the Fourier coefficients for  $Y_t(\tau)$ ;  $p_0, p_k, q_k$  are the Fourier coefficients for  $\mu(\tau)$ ;  $u_0, v_k, u_k$  are the coefficients for  $\beta(\tau)$ ;  $\eta_{t,0}, \omega_{t,k}, \eta_{t,k}$  are for  $\epsilon_t(\tau)$ .

We approximate the functional terms as  $Y_{t,m_t}(\tau), \mu_{m_t}(\tau), \beta_{m_t}(\tau), \epsilon_{t,m_t}(\tau)$ , respectively, where  $m_t$  is a hyperparameter for sieves, which controls the smoothing degree and balance the bias and variance of the approximation on the sieves.

The orthogonality property of complete Fourier bases creates the following relationship between the Fourier coefficients:

$$\begin{aligned}
 a_{t,0} &= p_0 + u_0 a_{t-1,0} + \eta_{t,0}, \\
 b_{t,k} &= p_k + \frac{1}{\sqrt{2}}(u_k b_{t-1,k} + \vartheta_k a_{t-1,k}) + \omega_{t,k}, \quad k = 1, \dots, m_t, \\
 a_{t,k} &= q_k + \frac{1}{\sqrt{2}}(u_k a_{t-1,k} - \vartheta_k b_{t-1,k}) + \eta_{t,k}, \quad k = 1, \dots, m_t.
 \end{aligned} \tag{2}$$

Let  $y_t \in \mathbb{R}^{K_t \times 1}$  be the Fourier coefficients of  $Y_{t,m_t}(\tau)$ , where  $y_t = (a_{t,0}, b_{t,1}, a_{t,1}, \dots, b_{t,m_t}, a_{t,m_t})^\top \in \mathbb{R}^{K_t \times 1}$  and  $K_t = 2m_t + 1$ . Similarly,  $e_t^\top \in \mathbb{R}^{K_t \times 1}$  be the Fourier coefficients of  $\epsilon_{t,m_t}(\tau)$ , where  $e_t = (\eta_{t,0}, \omega_{t,k}, \eta_{t,k}, \dots, \omega_{t,m_t}, \eta_{t,m_t})^\top \in \mathbb{R}^{K_t \times 1}$ .

After the projection, the FAR(1) model can be represented as:

$$y_t = \Gamma_0 + \Gamma^\top y_{t-1} + e_t, \tag{3}$$

where  $\Gamma_0 = (p_0, p_1, q_1, \dots, p_{m_t}, q_{m_t})^\top \in \mathbb{R}^{K_t \times 1}$  is an unknown vector of intercept coefficients and  $\Gamma \in \mathbb{R}^{K_t \times K_t}$  represents the unknown Fourier coefficients of the AR operator  $\beta_{m_t}(\tau)$ :

$$\Gamma = \begin{pmatrix} u_0 & & & & & & \\ & \frac{1}{\sqrt{2}}u_1 & -\frac{1}{\sqrt{2}}\vartheta_1 & & & & \\ & \frac{1}{\sqrt{2}}\vartheta_1 & \frac{1}{\sqrt{2}}u_1 & & & & \\ & & & \ddots & & & \\ & & & & \frac{1}{\sqrt{2}}u_{m_t} & -\frac{1}{\sqrt{2}}\vartheta_{m_t} & \\ & & & & \frac{1}{\sqrt{2}}\vartheta_{m_t} & \frac{1}{\sqrt{2}}u_{m_t} & \\ & & & & & & \end{pmatrix}_{K_t \times K_t}$$

## 2.2 Estimation with Mathematical Programming

For an optimal solution, we estimate the model parameters with a Mixed Integer Programming (MIP) approach. The FAR model is estimated by minimizing the absolute sum of errors:

$$\min \sum_{t=2}^T \mathbf{1}_{K_t}^\top |e_t|$$

where  $e_t = y_t - \Gamma_0 - \Gamma^\top y_{t-1}$ . In order to linearize the objective function, we rewrite the absolute sum of the error as the sum of two non-negative vectors  $e_t^+$  and  $e_t^-$ , see [4]. Then, we estimate the unknown model parameters by minimizing

total sum of the variables in the two non-negative vectors as follows:

$$\begin{aligned}
& \min \sum_{t=2}^T \mathbf{1}_{K_t}^\top (\mathbf{e}_t^+ + \mathbf{e}_t^-) \\
& \text{s.t. } \mathbf{y}_t - \Gamma_0 - \Gamma^\top \mathbf{y}_{t-1} = \mathbf{e}_t^+ - \mathbf{e}_t^- \quad \text{for } t = 2, \dots, T \\
& \quad \Gamma_{j,j} = \Gamma_{j+1,j+1}, \text{ where } j = 2, 4, \dots, K_t - 1, \\
& \quad \Gamma_{j+1,j} = -\Gamma_{j,j+1}, \text{ where } j = 2, 4, \dots, K_t - 1, \\
& \quad \Gamma_{j+i,j+i} = 0, \text{ where } j = 2, 4, \dots, K_t - 1, \text{ and } 2 \leq i \leq K_t, \\
& \quad \Gamma_{j,i} = 0, \text{ where } j < i, \quad j = 1, 3, \dots, K_t, \\
& \quad \mathbf{e}_t^+, \mathbf{e}_t^- \in \mathbb{R}_{\geq 0}^{K_t \times 1}, \quad \text{for } t = 2, \dots, T.
\end{aligned}$$

### 3 Real data analysis

#### 3.1 Data and Forecast setup

We next illustrate our proposed method with a real example. The data are collected from a large-scale German gas transmission network. Our dataset contains hourly gas in-flow and out-flow for 1029 gas nodes, including demand, supply, and storage nodes. The hourly gas in-flow or out-flow is observed hourly for the consecutive  $T=637$  days from April 19 Y1 to January 16 Y3. Y1 denotes the first year in the time series. The response (i.e.,  $Y_t(\tau)$ ) considered here is the daily gas flow curve at day  $t$ , see Fig. 1. After omitting three nodes with all-zero flows in the period of Y1 and Y2, we consider 1026 gas nodes for the forecast procedure.

The dataset is split into training and testing samples. Our objective is to predict a day-ahead hourly gas flow in the network based on one-day lagged historical curves of network gas flow. The training sample covers 365 days from April 19, Y1 to April 18, Y2 (Labeled as T1, total of 365 days). In T1, we optimize the model to obtain optimal parameter estimates. The rest of the total period from April 19, Y2 to January 16, Y3, is left for testing the model (labeled as T2, a total of 272 days). We select  $K_t = 15$ , and with a rolling window size of 365 days, we move forward one day at a time to update the parameter matrix, and we then forecast until we reach the end of the sample. In the testing period, the forecast performance is evaluated using an out-of-sample day-ahead hourly forecast of gas demand and supply.

#### 3.2 Evaluation

We measure forecast accuracy in terms of out-of-sample prediction errors. We convert back to the original gas flow data to facilitate comparison when computing the forecast accuracy. Specifically, we calculate the mean absolute error (MAE) and the mean absolute percentage error (MAPE) for hourly day-ahead

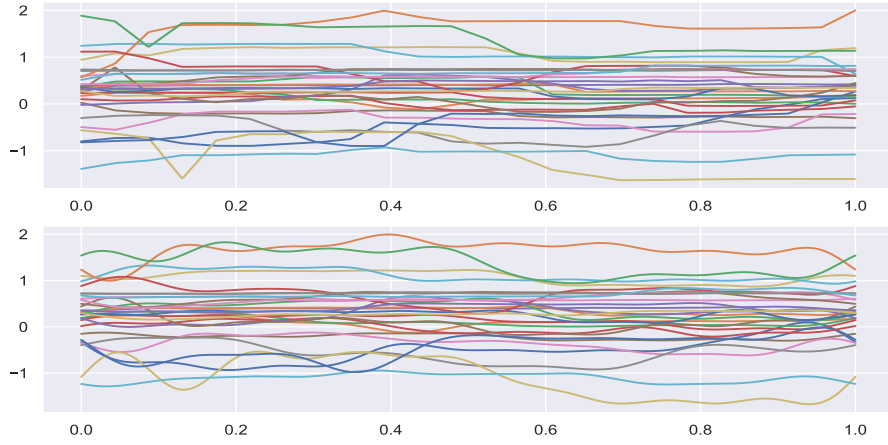


Fig. 1: Fourier series expansion of period one and #Basis=15 to represent a daily gas flow  $\{Y_t(\tau)\}_{t=1}^{30}$  at the largest gas node for one month. The time series values are normalized.

forecast at each node  $n, n = 1, \dots, N$ , as

$$\text{MAE}_{h,n} = \frac{1}{T_2} \sum_{t \in T_2} |Y_{t,h,n} - \hat{Y}_{t,h,n}|, \quad \text{MAPE}_{h,n} = \frac{1}{T_2} \sum_{t \in T_2} \left| \frac{Y_{t,h,n} - \hat{Y}_{t,h,n}}{Y_{t,h,n}} \right|,$$

where interval  $T_2$  indicates the total number of days in the test period.  $\hat{Y}_{t,h,n}$  is the  $h$ -hours ahead forecast of gas flow at day  $t$  at node  $n$ .

### 3.3 Forecast accuracy

We demonstrate the day-ahead hourly out-of-sample forecasting results in the gas network. Table 1 shows the median of the forecast errors, including MAE and MAPE over 1026 gas nodes by the proposed method and the benchmark alternative prediction methods. The considered alternative prediction models include the Autoregressive model of lag 1 and Long short-term memory (LSTM) model, representing deep learning model. LSTM model is built with 1 layer, 128 neurons, dropout is 0.1 and for every node it was trained with 100 epochs. In each column, the highlighted values indicate the best-performing result. Both median MAE and median MAPE show that the FAR model outperforms the AR and LSTM models for most of the hours in prediction for most nodes in the network. The weak prediction accuracy of the FAR model in the beginning (first 4 hours) and at the end of the forecast horizon (last 1-2 hours) can be explained by the not accurate smoothing of the original time series by the Fourier series as can be seen in Fig. 1, which leaves a room for further studies to enhance the proposed model.

Hour	MAE			MAPE		
	LSTM	AR(1)	FAR(1)	LSTM	AR(1)	FAR(1)
1	2.56	<b>0.11</b>	2.62	0.41	<b>0.02</b>	0.36
2	2.29	<b>0.72</b>	2.04	0.40	<b>0.14</b>	0.31
3	2.20	<b>1.12</b>	1.89	0.42	<b>0.22</b>	0.32
4	2.18	<b>1.45</b>	1.98	0.43	<b>0.28</b>	0.31
5	2.16	<b>1.69</b>	2.16	0.45	<b>0.35</b>	<b>0.35</b>
6	2.12	<b>1.88</b>	2.12	0.46	0.41	<b>0.35</b>
7	2.09	<b>2.05</b>	2.09	0.44	0.45	<b>0.36</b>
8	2.08	2.19	<b>2.00</b>	0.48	0.53	<b>0.36</b>
9	2.07	2.31	<b>1.88</b>	0.48	0.59	<b>0.36</b>
10	1.99	2.34	<b>1.85</b>	0.46	0.60	<b>0.36</b>
11	1.96	2.31	<b>1.83</b>	0.46	0.60	<b>0.35</b>
12	1.97	2.31	<b>1.79</b>	0.43	0.57	<b>0.34</b>
13	1.98	2.26	<b>1.75</b>	0.44	0.54	<b>0.33</b>
14	1.95	2.24	<b>1.76</b>	0.40	0.50	<b>0.32</b>
15	1.90	2.25	<b>1.75</b>	0.39	0.52	<b>0.32</b>
16	1.81	2.48	<b>1.72</b>	0.40	0.57	<b>0.33</b>
17	1.77	3.34	<b>1.60</b>	0.47	0.86	<b>0.38</b>
18	1.84	3.93	<b>1.58</b>	0.59	1.20	<b>0.49</b>
19	1.88	4.05	<b>1.65</b>	0.63	1.32	<b>0.59</b>
20	1.89	4.05	<b>1.74</b>	0.58	1.20	<b>0.56</b>
21	1.89	3.99	<b>1.85</b>	<b>0.57</b>	1.21	<b>0.57</b>
22	<b>1.87</b>	3.82	1.94	<b>0.54</b>	1.12	0.60
23	<b>1.75</b>	3.20	2.30	<b>0.46</b>	0.81	0.58
24	<b>2.20</b>	2.29	2.66	<b>0.39</b>	0.42	0.43

Table 1: Day-ahead hourly forecast: median MAE and MAPE of gas flow prediction at 1026 individual gas nodes from April 19, Y2, to January 16, Y3. The best performance for each hour is highlighted with **bold**.

## 4 Conclusion

We propose solving the FAR model with a mathematical programming approach for predicting day-ahead hourly time series. The real data analysis results show that the FAR model is more accurate for long-term high-frequency forecasts than the benchmark time series models.

**Acknowledgements** The work for this article has been supported by the European Union’s Horizon 2021 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101064994, and in collaboration with the Research Campus Modal funded by the German Federal Ministry of Education and Research (fund numbers 05M14ZAM, 05M20ZBM).

## References

1. Bertsimas, D., King, A., Mazumder, R.: Best subset selection via a modern optimization lens. *Ann. Statist.* 44 (2) 813 - 852 (2016). <https://doi.org/10.1214/15-AOS1388>



2. Chen, Y., Koch, T., Lim, K. G., Xu, X., and Zakiyeva, N.: A review study of functional autoregressive models with application to energy forecasting. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(3):e1525 (2021). [10.1002/wics.1525](https://doi.org/10.1002/wics.1525)
3. Chen, Y. and Li, B.: An adaptive functional autoregressive forecast model to predict electricity price curves. *Journal of Business & Economic Statistics*, 35(3):371–388 (2017). [10.1080/07350015.2015.1092976](https://doi.org/10.1080/07350015.2015.1092976)
4. Petkovic, M., Chen, Y., Gamrath, I., Gotzes, U., Hadjidimitrou, N. S., Zittel, J., Xu, X., and Koch, T.: A hybrid approach for high precision prediction of gas flows. *Energy Systems*, 13(2):383–408 (2022). [10.1007/s12667-021-00466-4](https://doi.org/10.1007/s12667-021-00466-4)
5. Xu, X., Chen, Y., Zhang, G., and Koch, T.: Regularized partially functional autoregressive model. Available at SSRN 3482262. (2020).