NONLINEAR NETWORK AUTOREGRESSIVE MODEL WITH APPLICATION TO NATURAL GAS NETWORK FORECASTING

XIAOFEI XU
RISK MANAGEMENT INSTITUTE, NATIONAL UNIVERSITY OF SINGAPORE, SINGAPORE,
E-MAIL: RMIXUX@NUS.EDU.SG, ORCID: 0000-0002-8019-9245
NAZGUL ZAKIYEVA
DEPARTMENT OF STATISTICS & APPLIED PROBABILITY, NATIONAL UNIVERSITY OF SINGAPORE, SINGAPORE,
MATHEMATICAL OPTIMIZATION DEPARTMENT, ZUSE INSTITUTE BERLIN, GERMANY,
E-MAIL: NAZGUL.ZAKIYEVA@U.NUS.EDU, ORCID: 0000-0001-9106-9916

Received June 22, 2020; revised June 24, 2020

Abstract. We propose a nonlinear network autoregressive (NNAR) model to investigate the dynamics of complex network time series with high-dimensionality and nonlinear spatial-temporal dependence. We assume that the current network at a given time point non-linearly depends on the lagged values, neighborhood effect, and a set of node-specific covariates via a nonparametric smooth function. We conduct estimation using the profile least square method where the unknown link function is estimated using the local linear regression technique. We demonstrate the application of the NNAR with the daily natural gas flows in a real-life high-pressure gas pipeline network, where the response is the high dimensional vector of gas flows at 128 nodes. The NNAR model provides more accurate forecasts of the gas flow network compared to the linear network vector autoregression model proposed by Zhu et al. (2017) and some multivariate autoregression and naive benchmark models.

1. Introduction

In this data-rich era, the development in data acquisition and storage has made it available to collect large-scale network data in many fields varying from biomedical sciences (Wu et al. 2014) and physics (Benson et al. 2016) to finance (Chen et al. 2018; Zhu et al. 2019) and socialization (Wasserman & Faust 1994; Zhu et al. 2017). Network data contains rich information for statistical inference, while the complexity of data with high-dimensionality, spatial-temporal dependence structure, non-linearity, and dynamic evolution creates extra challenges for statistical modeling and computation. To describe the dynamics and make a prediction of this complex data effectively and efficiently, it requires more flexible time series modeling in addition to the conventional tools that are designed for linear and low-dimensional time series data.

1991 *Mathematics Subject Classification.* subject classifications .
Key words and phrases. Network data; Nonlinear vector autoregression; Natural gas flows; Prediction.

Let Y_{it} be the continuous response collected from subject (node) i at time point t with $0 < t \leq T$ (e.g., natural gas flows at certain location). Accordingly, denote $\mathbf{Y}_t = (Y_{1t}, ..., Y_{Nt}) \in \mathbb{R}^N$ as an ultra-high dimensional vector with a large number of total nodes N. We assume that Y_{it} exhibits serial dependence on previous values and has certain spatial correlation among the nodes, i.e., network structure, and the dependence is unnecessary to be linear. In the literature of time series, some common univariate models such as AutoRegressive (AR) and AutoRegressive Moving Average (ARMA) based models have been well studied to forecast the dynamics of serially dependent time series in both theory and applications. However, these models study each time series Y_{it} separately, and the rich correlation information across different time series, e.g. the lead-lag dependence, is lost. Multivariate models such as Vector Autoregression (VAR) (Lütkepohl 2005; Box et al. 2015) are proposed with the information of multiple time series fully considered. On the other hand, the linear relationship assumption may not be valid in practice. Many nonlinear models have been proposed, such as functional coefficient autoregressive models for univariate time series (Huang & Shen 2004; Cai et al. 2000) and nonlinear VAR models for multivariate time series (Härdle et al. 1998); See also Fan & Yao (2008) for a good reference. However, in a high-dimensional case with sufficiently large N, these linear and nonlinear multivariate regression models will suffer from the "curse of dimensionality", which will cause a problem of overfitting and overparameterization, leading to poor out-of-sample forecast accuracy as well. Thus it is necessary to reduce dimension and many techniques have been therefore proposed, see e.g. factor modeling (Pan & Yao 2008; Park et al. 2009) and penalty estimation (Hsu et al. 2008). However, besides the serial and cross dependence, there could exist network information among the N nodes which needs to be taken into consideration.

The recent development of network modeling provides a wide variety of tools and methods to model the high-dimensional and complexly structured time series. In the literature, researchers mainly focus on conducting static analysis of network structure (Lee et al. 2010; Chen et al. 2013; Zhao et al. 2012; Zhou et al. 2017) or investigating the inherent dynamics of the network over time (Chen et al. 2018; Zhu & Pan 2018; Zhu et al. 2019). Among others, Zhu et al. (2017) proposed the linear network vector autoregression (NAR) model to study the dynamics of large scale network with the network information among individuals

incorporated, while dependence between the response and explanatory variables is assumed to be linear. Though convenient, in reality, linear regression-based models may not capture the data-driven complex relationship with both high-dimensionality and non-linearity. The single-index model (Carroll et al. 1997) is a prevalent way to flexibly handle the data-driven non-linearity and circumvent the problem of high-dimensionality simultaneously. For example, Jia et al. (2019) proposed partial autoregression single-index (PASI) model to handle linear network dependence and nonlinear influence of static covariates; See also Wang & Yang (2009), Yu & Ruppert (2002), Li & Genton (2009), and Liang et al. (2010) for more reference of single-index model. However, Jia et al. (2019) did not consider the non-linearity of network dependence. In our work, we propose a flexible semiparametric model also by combining the single-index technique and dynamic autoregression network model, which inherits the advantages of both models and handles the non-linearity of network dependence as well.

We propose a nonlinear network autoregressive (NNAR) model to investigate the dynamics of network time series with nonlinear spatial-temporal dependence in high-dimensional framework, and simultaneously allow the measurement of the nonlinear impact of multivariate node-specific covariates, if applicable. In particular, we assume response Y_{it} depends on a single index defined on three items: its own lagged value, the weighted average of its neighbors, and exogenous covariates via a nonlinear link function, referring as the momentum effect, the network effect and the nodal effect by Zhu et al. (2017), respectively. The link function is assumed to be unknown and smooth. We conduct estimation using the profile least square method (Fan & Gijbels 1996) where the link function is estimated using the local linear regression technique. This paper makes contributions in two aspects: (1) We propose a flexible nonlinear network autoregressive model to investigate the dynamics of large-scale network with complex spatial-temporal dependence. The proposed model helps capture the nonlinear network dependence and node-specific exogenous covariates' impact. While Zhu et al. (2017) and Jia et al. (2019) only considered the linear network dependence. (2) We demonstrate the application of the NNAR model on forecasting gas flows at 128 distribution nodes of a high-pressure natural gas transmission network in Europe. It

provides a more accurate out-of-sample forecast for the gas flows network compared with the linear network model and some multivariate time series and naive benchmark models.

The rest of the paper is organized as follows. Section 2 details the NNAR model and parameter estimation procedure using the profile least square method. Section 3 presents the gas flow network data on an energy transmission system. Section 4 implements the NNAR model to investigate the dynamics of gas flows network and conducts forecasting with comparison to several alternative models. Section 5 concludes.

2. Method

In this section, we present the nonlinear network autoregressive (NNAR) model and the estimation procedure using the profile least square method.

2.1. The NNAR model.

Recall that the number of nodes in the network is N, and Y_{it} is the continuous response collected from node i at time point t with $0 < t \leq T$ and $1 \leq i \leq N$. Denote $\mathbf{Y}_t = (Y_{1t}, ..., Y_{Nt}) \in \mathbb{R}^N$ an ultra-high dimensional vector with a large number of N. In addition, for each node i, assume a $p-$dimensional node-specific random vector $\mathbf{Z}_{it} = (Z_{it}^{(1)}, \ldots, Z_{it}^{(p)})^\top \in \mathbb{R}^p$ can be observed. To model $\mathbf{Y}_t$, we propose the following nonlinear network autoregressive model (NNAR):

$$Y_{it} = g\left(\beta_1 \sum_{j=1}^{N} w_{ij} Y_{j(t-1)} + \beta_2 Y_{i(t-1)} + \mathbf{Z}_{i(t-1)}^\top \boldsymbol{\gamma}\right) + \epsilon_{it}, \tag{1}$$

where $g(\cdot)$ is an unknown link function which is assumed be smooth. $w_{ij} \in [0, 1]$ is a given weight to measure the strength of the connection between node i and j for $i, j = 1, \cdots, N$. In specific, if we know the adjacency matrix of the network structure, which can be defined as $A = (a_{ij}) \in \mathbb{R}^{N \times N}$, where $a_{ij} = 1$ if there exists a relationship between node i and j, and $a_{ij} = 0$ otherwise, then w_{ij} is commonly defined as $w_{ij} = a_{ij}/n_i$ where $n_i = \sum_{j \neq i} a_{ij}$ is the total number of nodes that node i is connected, i.e. out-degree (Wasserman & Faust 1994). Such a choice is quite common in many kinds of research like graphical and social network analysis (Bondy & Murty 1976; Zhu et al. 2017). However, in our real data implementation to natural gas flow network at next section, the adjacency matrix for the gas network is unknown, we thus define w_{ij} as the inverse of the shortest path between the gas node i and j, with further located nodes given smaller weight. The quantity $\sum_{j=1}^{N} \omega_{ij} Y_{j(t-1)}$

characterizes the average impact from the network to ith node at time $t-1$. Its associated parameter β_1 is referred as the network effect. The term $Y_{i(t-1)}$ is the autoregressive term which stands for the serial dependence and β_2 is the corresponding parameter. The term $\mathbf{Z}_{i(t-1)}^{\top}\boldsymbol{\gamma}$ evaluates the influence of exogenous impact to the ith node at time $t-1$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^{\top} \in \mathbb{R}^p$ is the associated coefficient (i.e. exogenous nodal effect). Moreover, ϵ_{it} is the error term, we assume that it is independent to response and exogenous covariates, i.e., $E(\epsilon_{it}, Y_{is}) = 0$ and $E(\epsilon_{it}, Z_{is}^{(j)}) = 0$ for any $s < t$ and $j = 1, \cdots, p$, and follows normal distribution with $\epsilon_{it} \sim_{\text{i.i.d}} \text{N}(0, \sigma^2)$.

For notation simplicity, let $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\top} \in \mathbb{R}^2$, and $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\top}, \boldsymbol{\gamma}^{\top})^{\top} \in \mathbb{R}^{p+2}$ standing for the vector of all the unknown parameters. We further employ the profile least-square estimation technique to estimate $\boldsymbol{\theta}$ and the unknown link function $g(\cdot)$. We rewrite the NNAR model in (1) as

$$
\begin{aligned}
Y_{it} &= g\left(\beta_1 \sum_{j=1}^{N} w_{ij} Y_{j(t-1)} + \beta_2 Y_{i(t-1)} + \mathbf{Z}_{i(t-1)}^{\top}\boldsymbol{\gamma}\right) + \epsilon_{it}, \\
&= g\left(\mathbf{X}_{it}^{\top}\boldsymbol{\theta}\right) + \epsilon_{it},
\end{aligned} \tag{2}
$$

where $\mathbf{X}_{it} = (\sum_{j=1}^{N} w_{ij} Y_{j(t-1)}, Y_{i(t-1)}, \mathbf{Z}_{i(t-1)}^{\top})^{\top} \in \mathbb{R}^{p+2}$ is the variable vector consisting of the network effect, lag effect and exogenous nodal effect to node i at time $t-1$. To ensure the identification, we set $||\boldsymbol{\theta}||_2 = 1$ with the first element positive, where $||\cdot||_2$ is the L_2 norm.

2.2. **Estimation.** In semiparametric models, it is popular to estimate the unknown parametric components using the profile likelihood approach, see Liang et al. (2010). We apply this technique for estimating parameter $\boldsymbol{\theta}$. First we start estimating the nonlinear link function $g(\cdot)$ for a given parameter value $\boldsymbol{\theta}$ using the local linear approximation method (Fan & Gijbels 1996). In particular, by defining $u_{it} = \mathbf{X}_{it}^{\top}\boldsymbol{\theta} \in \mathbb{R}$ we linearly approximate the function $g(\cdot)$ at a given point u_0. Since the unknown function $g(\cdot)$ is assumed to be smooth, we approximate it locally by a linear function (Taylor expansion),

$$
g(u) \approx g(u_0) + g'(u_0)(u - u_0), \quad \text{for } u \in u_0 \pm h,
$$

where h is a bandwidth referring to the size of the neighborhoods that the linear approximation holds. For notation simplicity, denote $a = g(u_0)$, and $b = g'(u_0)$. This leads to the

following locally approximated NNAR model:

$$\begin{aligned} Y_{it} &= g(u_{it}) + \epsilon_{it}, \\ &\approx a + b(u_{it} - u_0) + \epsilon_{it},\ u_{it} \in u_0 \pm h. \end{aligned}$$

We consider to estimate the local parameters a and b by minimizing the below objective function using the weighted least squares method:

$$S(a,b) = \sum_{t=1}^{T}\sum_{i=1}^{N}\Big(Y_{it} - a - b(u_{it} - u_0)\Big)^2 K_h(u_{it} - u_0), \tag{3}$$

where $K_h(\cdot) = K(\cdot/h)/h$, and $K(\cdot)$ is a nonnegative unimodal kernel function. In our study, we use the Gaussian kernel function and select the optimal bandwidth via cross-validation (CV) method.

By computing the derivatives of $S(a,b)$ with respect to a and b, we find the minimizers of Eq.(3) for a given value of $\boldsymbol{\theta}$:

$$\hat{g}(u_0;\boldsymbol{\theta}) = \hat{a}|_{\boldsymbol{\theta}} = \frac{K_{20}(u_0,\boldsymbol{\theta})K_{01}(u_0,\boldsymbol{\theta}) - K_{10}(u_0,\boldsymbol{\theta})K_{11}(u_0,\boldsymbol{\theta})}{K_{00}(u_0,\boldsymbol{\theta})K_{20}(u_0,\boldsymbol{\theta}) - K_{10}^2(u_0,\boldsymbol{\theta})}, \tag{4}$$

where $K_{j\ell}(u,\boldsymbol{\theta}) = \sum_{t=1}^{T}\sum_{i=1}^{N}[K_h(\mathbf{X}_{it}^\top\boldsymbol{\theta} - u)](\mathbf{X}_{it}^\top\boldsymbol{\theta} - u)^j Y_{it}^\ell$ with exponents $j = 0, 1, 2$ and $\ell = 0, 1$. The proof is given in the appendix.

Next we estimate the parameter $\boldsymbol{\theta}$ using the estimates of the nonparametric component $\hat{g}(u_0;\boldsymbol{\theta})$ in Eq. (4). If assuming $u_0 = \mathbf{X}_{it}^\top\boldsymbol{\theta}$, we have $Y_{it} \approx g(\mathbf{X}_{it}^\top\boldsymbol{\theta};\boldsymbol{\theta}) + \epsilon_{it}$. We then obtain the estimator $\hat{\boldsymbol{\theta}}$ by minimizing the following profile least-square function following the assumption of Jennrich (1969):

$$Q(\boldsymbol{\theta}) = \sum_{t=1}^{T}\sum_{i=1}^{N}\Big\{Y_{it} - \hat{g}(\mathbf{X}_{it}^\top\boldsymbol{\theta};\boldsymbol{\theta})\Big\}^2. \tag{5}$$

Since there is no closed-form solution for the estimate $\boldsymbol{\theta}$ in above Eq.(5), we apply stochastic gradient descent algorithm (Kushner & Yin 2003) to iteratively update the estimations by minimizing the above objective function. The calculation stops when the parameters converge, and we set the final iterative estimator as $\hat{\boldsymbol{\theta}}$.

Finally, plugging in the optimal value $\hat{\boldsymbol{\theta}}$ to Eq.(4), and replacing the notation of u_0 with a general parameter symbol u, we have the following estimate of link function $g(\cdot)$

$$\hat{g}(u;\hat{\boldsymbol{\theta}}) = \hat{a}|_{\hat{\boldsymbol{\theta}}} = \frac{K_{20}(u,\hat{\boldsymbol{\theta}})K_{01}(u,\hat{\boldsymbol{\theta}}) - K_{10}(u,\hat{\boldsymbol{\theta}})K_{11}(u,\hat{\boldsymbol{\theta}})}{K_{00}(u,\hat{\boldsymbol{\theta}})K_{20}(u,\hat{\boldsymbol{\theta}}) - K_{10}^2(u,\hat{\boldsymbol{\theta}})}, \tag{6}$$

where $K_{j\ell}(u,\boldsymbol{\theta})$ has the same definition as before.

3. Data

Natural gas has become an important and clean energy source for power systems with its advantages varying from lower pollutant emission and smaller construction period, to higher efficiency of conversion and loading. About 24% of the worldwide energy demand is met by natural gas in 2018 (BP 2019). In particular, natural gas is a key energy resource for Europe and accounts for about 20% of the European energy demand (Petkovic et al. 2019). Natural gas is transported through transit countries and to the local distribution nodes through the high-pressure transmission pipeline network which is operated by so-called transmissions system operators or TSOs. The European gas market is moving to more short-term operations, for example, day-ahead contracts. This increases the necessity of modeling the underlying network dynamics of future gas flows for not only one node or a few nodes, but large-dimensional nodes in the transmission network. Accurate short-term forecasting of natural gas demand and supply is of importance for TSOs to monitor the situation and conduct operational decisions to ensure the safety of supply. There exists rich literature in natural gas forecasting, see e.g. Stoll & Wiebauer (2010), Soldo (2012), Banda & Herty (2008), Koch et al. (2015), Chen et al. (2018), and Chen et al. (2020). However, the dynamics of gas flow network structure has less been explored in the context of forecasting.

Our work is motivated by the challenging problem of short-term forecasting of gas supply and demand in the high-dimensional gas transmission network. We collect the high-resolution natural gas flow data at N=128 nodes in the gas pipeline network in one European country. The daily average gas in-flow or out-flow is observed for the consecutive T=637 days over 22 months. The gas flow network data is standardized with zero mean and unit variance to use because of the significant scale difference of flow values at various nodes. The response (i.e. Y_{it}) considered here is the daily average gas flow of node $i \in \{1, \cdots, 128\}$ at day $t \in [1, T]$. In addition, we consider the daily average air temperature at each node as an exogenous variable. Given that the natural gas is being widely used for heating purposes in European countries, the temperature is usually considered as a possible affected factor in forecasting gas flows (Chen et al. 2018). As gas flows in/out through all nodes in the network where gas nodes are connected with a pipeline, and it is unclear about the adjacency matrix $A = (a_{ij})$ of the network as well, we define the weight matrix, denoted as

W with the $(i,j)-$th element being the weight w_{ij}, as the inverse shortest path among each pair of nodes i and j. Here, the shortest path is defined as the Euclidean distance between two nodes.

Figure 1 displays the time series plot of gas flows at 25 arbitrarily selected nodes as a graphical demonstration. We can see similar dynamics of gas flow time series at these nodes, and synchronous behaviors associating with seasons. Figure 2 displays the lag-1 sample cross-correlation matrix of natural gas flows at 128 nodes. We can see a strong correlation among the 128 nodes. Figure 3 displays the correlation coefficients between variables from day 1 to day 637. In the figure, the columns represent the response variable of gas flow Y_t, network term WY_{t-1} referring to the network impact, lag-1 gas flow variable Y_{t-1} representing momentum impact, and the temperature variable Z_{t-1} representing node-specific exogenous covariate's impact from left to right respectively. As shown, the Pearson linear correlation coefficients between Y_t and three regressive terms are 0.48, 0.84, and -0.59 respectively, which indicate the existance of correlation among them. In addition, we can see some nonlinear dependence from the scatter plots. These motivate the use of the NNAR model which helps jointly analyze the nonlinearity, momentum impact, network impact, and exogenous variable effect simultaneously to utilize the rich information.

4. Forecasting Results

In this section, we demonstrate the forecasting performance of the NNAR model using the natural gas network data described in Section 3. We perform out-of-sample forecasts of daily natural gas and compare it with several alternative methods.

4.1. **Setup and evaluation.**

We divide the network dataset into two phases with the first 500 days used as training period (T_1), which covers 80% of the total period, and 137 days from day 501 to the end at day 637 as the forecasting period (T_2). We train the model and estimate parameters in T_1, and select the optimal bandwidth as $h = 4$ via cross-validation.

As alternative methods, we consider the linear NAR model (Zhu et al. 2017), VAR and naive methods including Random Walk and Sample Mean to forecast 1-day ahead daily natural gas flows of 128 nodes. The NAR model considers linear network dependence among nodes. VAR method is popular in forecasting gas consumption, and we select the

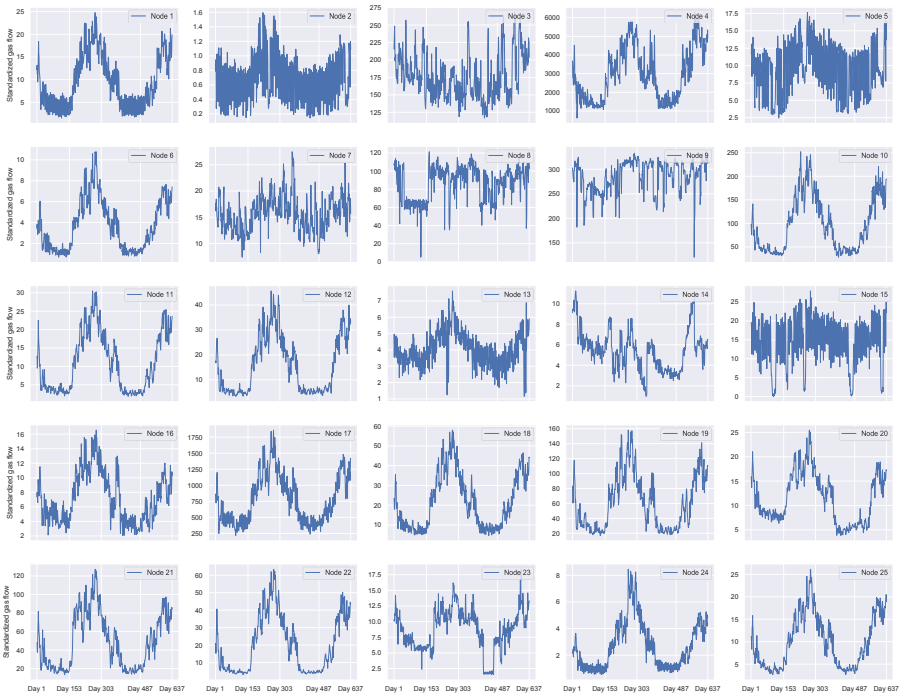

FIGURE 1. Time series plot of gas flows at 25 illustrative nodes from day 1 to day 637.

lag order of VAR via Bayesian Information Criteria (BIC). In addition, naive forecasting methods of Sample Mean (SM) and Random Walk (RW) are used as benchmark models. Here, at time t, the SM model forecasts h-step-ahead value by taking the average of all observed data up to time t, that is, $\hat{Y}_{i,t+h} = \frac{1}{t}\sum_{j=1}^{t} Y_{i,j}$, for $i = 1, \ldots, N$.

We evaluate the relative forecast accuracy according to the average forecast error of individual nodes in the network. We use mean absolute percentage error (MAPE) as an error evaluation criteria. The smaller the MAPE, the better accuracy is obtained by the forecast model. First, for each node $i = 1 \ldots N$ we obtain daily predicted gas flow series $\hat{Y}_{it}$. The 1-day-ahead prediction performance is evaluated over the forecasting period of T_2 for each node. The MAPE for each node i is obtained as:

$$\text{MAPE}_i = \frac{1}{|T_2|}\sum_{t\in T_2}\left|\frac{\hat{Y}_{it} - Y_{it}}{Y_{it}}\right|,$$

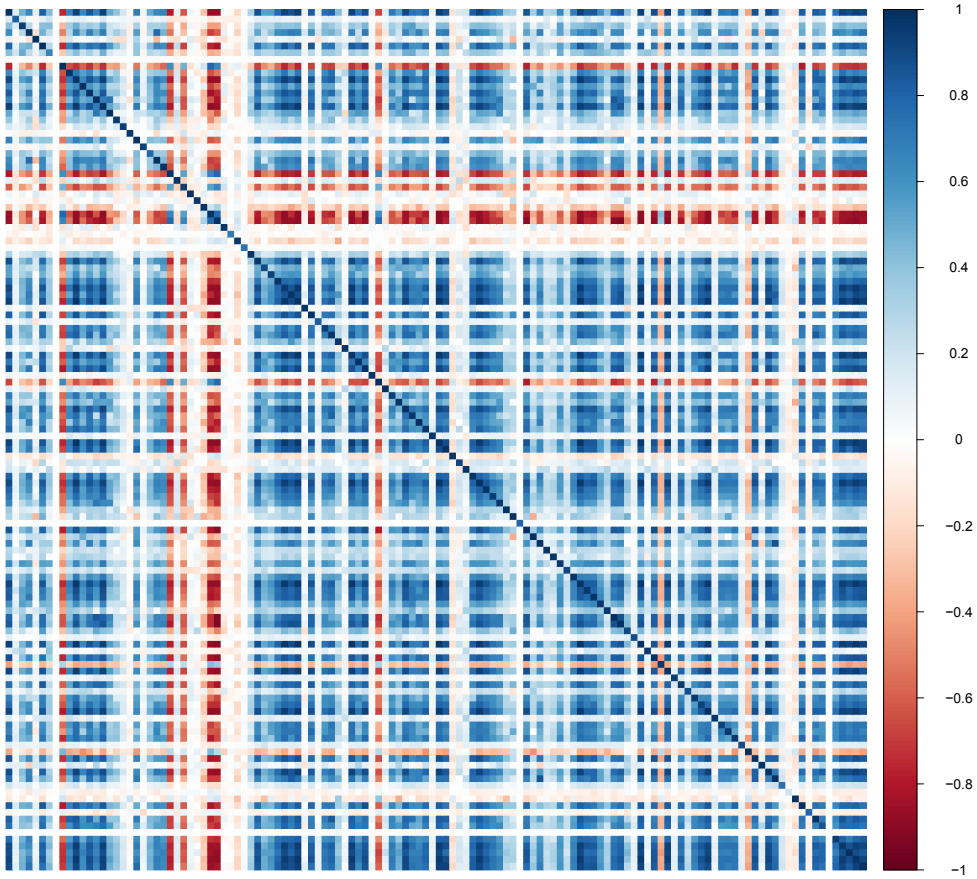

FIGURE 2. Sample lag 1 cross-correlation matrix of 128 nodes.

for $i = 1, \cdots, 128$, where $|T_2|$ is the length of forecasting period T_2. Average forecast performance evaluation is obtained via an average MAPE as

$$\text{aMAPE} = \frac{\sum_{i=1}^{N} \text{MAPE}_i}{N}.$$

4.2. **Results.** We demonstrate the 1-day-ahead out-of-sample forecasting results in the large scale gas network. Table 1 reports the aMAPE and its standard deviation (sd) as well as range over 128 nodes of de-standardized gas-flows for the NNAR model and the alternative models. The MAPE over different models is compared in the boxplot of Figure 4. As can be seen, the NNAR model performs much better than the VAR model, the RW, and the SM models with smaller aMAPE, smaller sd, and more narrow range. The NNAR model slightly outperforms the NAR model with aMAPE of 13.06%, sd of 11.589%, and

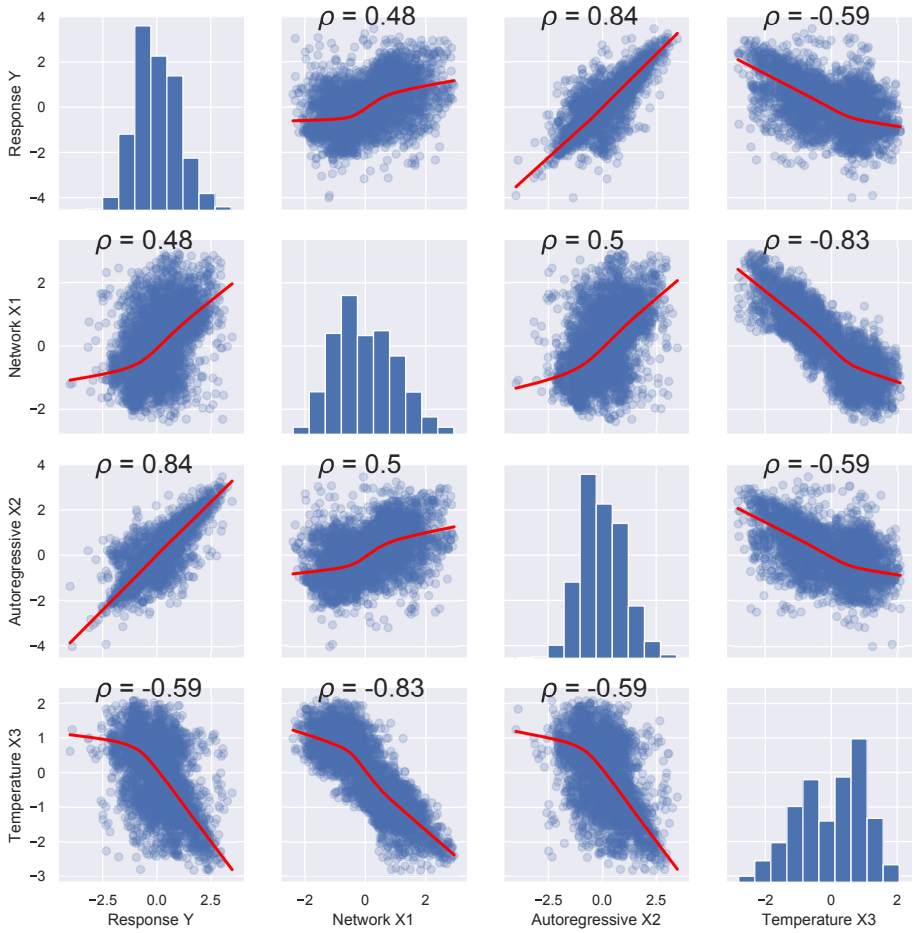

FIGURE 3. Correlation coefficients between variables from day 1 to day 637. The columns represent the response variable of gas flow Y_t, network term WY_{t-1} (the network impact), lag-1 gas flow Y_{t-1} (momentum impact), and exogenous variable Z_{t-1}(temperature) from left to right respectively. Here, ρ is the Pearson correlation coefficient.

range [2.206%; 84.452%] respectively. Our model is only slightly better than the linear NAR model, this could be because the non-linearity dependence in the gas network is not that significant, and the linear network model can fit the gas date well. It is worth mentioning that the NNAR model provides some advantage in flexibly capturing real network data with either linear or nonlinear dependence structure compared to the linear NAR model.

	aMAPE	(sd)	Range
NNAR	13.060%	(11.589%)	[2.206%; 84.452%]
NAR	13.570%	(12.398%)	[2.104%; 88.598%]
Random Walk	47.405%	(19.268%)	[3.817%; 98.546%]
VAR (BIC)	38.405%	(20.324%)	[3.878%; 149.517%]
Sample Mean	38.405%	(20.324%)	[3.878%; 149.517%]

TABLE 1. The 1-day-ahead out-of-sample forecast performance of the NNAR and alternative models. Average MAPE (aMAPE) and its standard deviation (sd) as well as range (Range) over 128 nodes are reported.

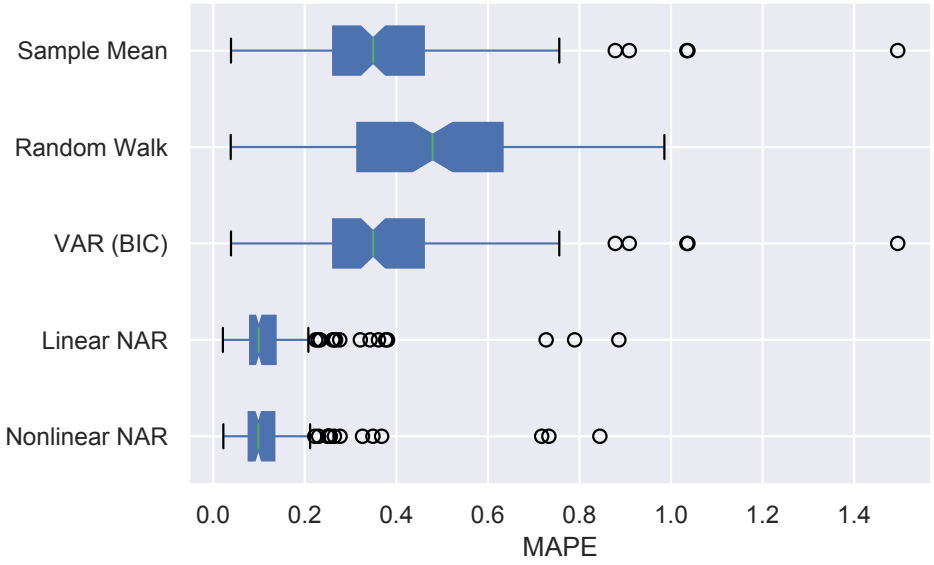

FIGURE 4. The boxplot of MAPE of the NNAR and four alternative models over 128 nodes.

Figure 5 displays the 1-day-ahead out-of-sample forecasts of the NNAR model for daily gas flow at the 25 nodes as an illustration from day 501 to the end at day 637. We can see that the NNAR model successfully captures the dynamic evolution of the gas flow time series at each individual node. The NNAR model also fits the observed gas flows well at the rest nodes. In general, we find that our proposed model delivers stable forecast performance no matter the node's type and dynamic. We have to mention that since we analyze the gas network partially i.e. with a selected number of gas nodes, and the nodes connection information in pipeline network is not clear as well, the conclusion from our modeling may have some limitations.

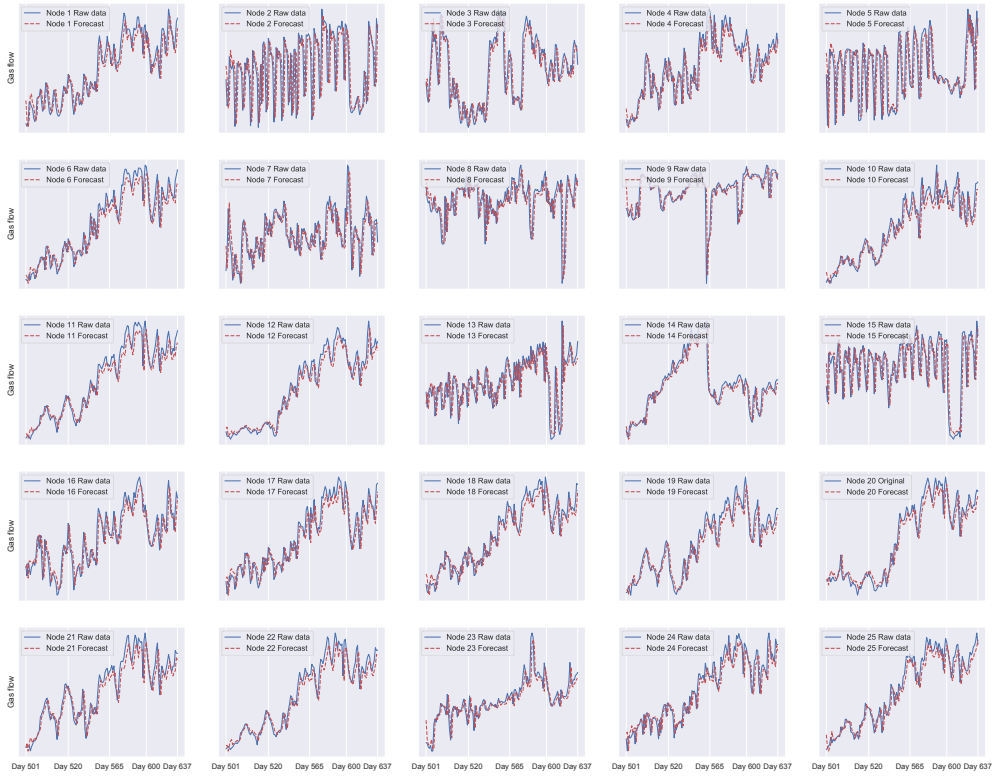

FIGURE 5. The daily gas flows and 1-step-ahead forecast of the NNAR model at 25 nodes from day 501 to day 637. The data are normalized to display.

5. CONCLUSION

We propose a nonlinear network autoregressive model to investigate the complex dynamics of high-dimensional network with nonlinear spatial-temporal dependence structure, where the nonlinear impact of node-specific exogenous covariates is incorporated simultaneously. The proposed model assumes that the current network at a given time point non-linearly depends on three items: the past values, network effect, and exogenous covariates via a nonlinear smooth function. We conduct estimation using the profile least square method where the unknown link function is estimated via the local linear regression technique. We demonstrate the application of the NNAR with the daily natural gas flows in a real-life high-pressure gas pipeline network, where the response is the high dimensional

vector of gas flows at 128 nodes. The NNAR model provides more accurate forecasts of the gas flow network compared with several alternative models. It shows that the NNAR model has some advantages in flexibly capturing real-life network data with either linear or nonlinear dependence structure compared to linear models.

Acknowledgements

We thank the Editor and anonymous referees for their valuable time and careful comments, which have helped improve this paper.

References

Banda, M. K. & Herty, M. (2008). Multiscale modeling for gas flow in pipe networks. *Mathematical Methods in the Applied Sciences*, *31*(8), 915–936.

Benson, A. R., Gleich, D. F., & Leskovec, J. (2016). Higher-order organization of complex networks. *Science*, *353*(6295), 163–166.

Bondy, J. A. & Murty, U. S. R. (1976). *Graph theory with applications*, volume 290. Macmillan London.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

BP, B. P. (2019). Statistical review of world energy 2019. *British Petroleum World Energy*.

Cai, Z., Fan, J., & Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, *95*(451), 941–956.

Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*(438), 477–489.

Chen, X., Chen, Y., & Xiao, P. (2013). The impact of sampling and network topology on the estimation of social intercorrelations. *Journal of Marketing Research*, *50*(1), 95–110.

Chen, Y., Chua, W. S., & Koch, T. (2018). Forecasting day-ahead high-resolution natural-gas demand and supply in Germany. *Applied Energy*, *228*, 1091–1110.

Chen, Y., Trimborn, S., & Zhang, J. (2018). Discover regional and size effects in global bitcoin blockchain via sparse-group network autoregressive modeling. *Available at SSRN 3245031*.

Chen, Y., Xu, X., & Koch, T. (2020). Day-ahead high-resolution forecasting of natural gas demand and supply in Germany with a hybrid model. *Applied Energy*, *262*, 114486.

Fan, J. & Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press.

Fan, J. & Yao, Q. (2008). *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.

Härdle, W., Tsybakov, A., & Yang, L. (1998). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference*, *68*(2), 221–245.

Hsu, N.-J., Hung, H.-L., & Chang, Y.-M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, *52*(7), 3645–3657.

Huang, J. Z. & Shen, H. (2004). Functional coefficient regression models for non-linear time series: a polynomial spline approach. *Scandinavian Journal of Statistics*, *31*(4), 515–534.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, *40*(2), 633–643.

Jia, Y., Song, T., Wu, S., Zhang, Q., & Su, Y. (2019). Dynamic influence prediction of social network based on partial autoregression single index model. *Discrete Dynamics in Nature and Society*, *2019*.

Koch, T., Hiller, B., Pfetsch, M., & Schewe, L. (2015). *Evaluating Gas Network Capacities*. SIAM.

Kushner, H. & Yin, G. G. (2003). *Stochastic approximation and recursive algorithms and applications*, volume 35. Springer Science & Business Media.

Lee, L.-f., Liu, X., & Lin, X. (2010). Specification and estimation of social interaction models with network structures. *The Econometrics Journal*, *13*(2), 145–176.

Li, Y. & Genton, M. G. (2009). Single-index additive vector autoregressive time series models. *Scandinavian Journal of Statistics*, *36*(3), 369–388.

Liang, H., Liu, X., Li, R., & Tsai, C.-L. (2010). Estimation and testing for partially linear single-index models. *Annals of Statistics*, *38*(6), 3811–3836.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Pan, J. & Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, *95*(2), 365–379.

Park, B. U., Mammen, E., Härdle, W., & Borak, S. (2009). Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association*, *104*(485), 284–298.

Petkovic, M., Chen, Y., Gamrath, I., Gotzes, U., Hadjidimitriou, N. S., Zittel, J., Xu, X., & Koch, T. (2019). A hybrid approach for high precision prediction of gas flows. Technical Report 19-26, ZIB.

Soldo, B. (2012). Forecasting natural gas consumption. *Applied Energy*, *92*, 26–37.

Stoll, S.-O. & Wiebauer, K. (2010). A spot price model for natural gas considering temperature as an exogenous factor with applications. *The Journal of Energy Markets*, *3*(3), 113.

Wang, L. & Yang, L. (2009). Spline estimation of single-index models. *Statistica Sinica*, 765–783.

Wasserman, S. & Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

Wu, S., Shao, F., Sun, R., Sui, Y., Wang, Y., & Wang, J. (2014). Analysis of human genes with protein–protein interaction network for detecting disease genes. *Physica A: Statistical Mechanics and its Applications*, *398*, 217–228.

Yu, Y. & Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, *97*(460), 1042–1054.

Zhao, Y., Levina, E., & Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, *40*(4), 2266–2292.

Zhou, J., Tu, Y., Chen, Y., & Wang, H. (2017). Estimating spatial autocorrelation with sampled network data. *Journal of Business & Economic Statistics*, *35*(1), 130–138.

Zhu, X. & Pan, R. (2018). Grouped network vector autoregression. *Statistica Sinca*.

Zhu, X., Pan, R., Li, G., Liu, Y., & Wang, H. (2017). Network vector autoregression. *The Annals of Statistics*, *45*(3), 1096–1123.

Zhu, X., Wang, W., Wang, H., & Härdle, W. K. (2019). Network quantile autoregression. *Journal of Econometrics*, *212*(1), 345–358.

Communicated by *Masanobu Taniguchi*

Appendix: Estimation derivation

Here we describe the estimation of the nonlinear function g in more detail. First, we write the NNAR model for node i at time point t as follows,

$$Y_{it} = g(\mathbf{X}_{it}^{\top}\boldsymbol{\theta}) + \epsilon_{it},$$

and define $u_{it} = \mathbf{X}_{it}^{\top}\boldsymbol{\theta}$. Then we rewrite the model in terms of u_{it} as

$$Y_{it} = g(u_{it}) + \epsilon_{it}.$$

We assume unknown link function $g(\cdot)$ is second order differentiable. At a given point u_0, we consider to approximate the function $g(\cdot)$ by their first order Taylor's expansion with respect to u as

$$g(u) = g(u_0) + g'(u_0)(u - u_0), \text{ for } u \in u_0 \pm h,$$

where h is a bandwidth. Denote $a = g(u_0)$, and $b = g'(u_0)$, we can estimate the nonlinear function g by minimizing the following objective function using the weighted least square method,

$$S(a,b) = \sum_{t=1}^{T}\sum_{i=1}^{N}[Y_{it} - a - b(u_{it} - u_0)]^2 K_h(u_{it} - u_0). \tag{A.1}$$

To find the minimizer of local parameters a and b in (A.1), we take its derivatives with respect to a and b, respectively, then we have

$$\begin{aligned}
\frac{\partial S}{\partial a} &= \sum_{t=1}^{T}\sum_{i=1}^{N}[Y_{it} - a - b(u_{it} - u_0)]K_h(u_{it} - u_0), \\
\frac{\partial S}{\partial b} &= \sum_{t=1}^{T}\sum_{i=1}^{N}[Y_{it} - a - b(u_{it} - u_0)](u_{it} - u_0)K_h(u_{it} - u_0).
\end{aligned}$$

Setting the above two derivatives to zero, we obtain

$$\sum_{t=1}^{T}\sum_{i=1}^{N}[Y_{it} - a - b(u_{it} - u_0)]K_h(u_{it} - u_0) = 0, \tag{A.2}$$

and

$$\sum_{t=1}^{T}\sum_{i=1}^{N}[Y_{it} - a - b(u_{it} - u_0)](u_{it} - u_0)K_h(u_{it} - u_0) = 0. \tag{A.3}$$

From (A.2) we obtain the following estimate for b as

$$\hat{b} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{it}K_h(u_{it} - u_0) - a\sum_{t=1}^{T}\sum_{i=1}^{N} K_h(u_{it} - u_0)}{\sum_{t=1}^{T}\sum_{i=1}^{N}(u_{it} - u_0)K_h(u_{it} - u_0)}. \tag{A.4}$$

For notation simplicity, we denote $K_h(u_{it}-u_0)=K_h^\star$. Plug (A.4) into (A.3), we can derive the estimate of a as:

$$\hat{a}=\frac{\sum_{t=1}^{T}\sum_{i=1}^{N}\{K_h^\star\big[\sum_{t=1}^{T}\sum_{i=1}^{N}(u_{it}-u_0)^2K_h^\star-(u_{it}-u_0)\sum_{t=1}^{T}\sum_{i=1}^{N}(u_{it}-u_0)K_h^\star\big]\}Y_{it}}{\sum_{t=1}^{T}\sum_{i=1}^{N}\{K_h^\star\big[\sum_{t=1}^{T}\sum_{i=1}^{N}(u_{it}-u_0)^2K_h^\star-(u_{it}-u_0)\sum_{t=1}^{T}\sum_{i=1}^{N}(u_{it}-u_0)K_h^\star\big]\}}.$$

Recall that $g(u_0)=a$, we can directly get

$$\hat{g}(u_0;\boldsymbol{\theta})=\hat{a}|_{\boldsymbol{\theta}}=\frac{K_{20}(u_0,\boldsymbol{\theta})K_{01}(u_0,\boldsymbol{\theta})-K_{10}(u_0,\boldsymbol{\theta})K_{11}(u_0,\boldsymbol{\theta})}{K_{00}(u_0,\boldsymbol{\theta})K_{20}(u_0,\boldsymbol{\theta})-K_{10}^2(u_0,\boldsymbol{\theta})}, \tag{A.5}$$

where $K_{j\ell}(u_0,\boldsymbol{\theta})=\sum_{t=1}^{T}\sum_{i=1}^{N}K_h^\star(\mathbf{X}_{it}^\top\boldsymbol{\theta}-u_0)^jY_{it}^\ell$, with exponents $j=0,1,2$ and $\ell=0,1$.