Hybrid VARX-SVR and GSTARX-SVR for Forecasting Spatio-Temporal Data

Suhartono, Bahagiati Maghfiroh, Santi Puteri Rahayu

Abstract: Generalized Space-Time Autoregressive or GSTAR model is a special form of Vector Autoregressive or VAR model and commonly used for forecasting spatio-temporal data. The objective of this study is to propose hybrid spatio-temporal methods by applying Support Vector Regression or SVR as a nonlinear machine learning method in two representations model, i.e. as VAR or GSTAR with exogenous variables known as VARX or GSTARX, respectively. These two proposed hybrid methods are then known as VARX-SVR and GSTARX-SVR model. These models consist of two steps modelling, i.e. the first step is modelling of trend, seasonal, and calendar variation effects using time series regression, and the residual of this first step is modelled by VARX-SVR and GSTARX-SVR in the second step. Both simulation and real data about inflow and outflow currency in three location of Bank Indonesia at West Java region are used as case studies. The results of simulation study show that both the proposed VARX-SVR and GSTARX-SVR models yield more accurate forecast in testing dataset than VARX and GSTARX. Furthermore, the results of real data showed that VARX is the best model for forecasting outflow in three locations and inflow in two locations. Meanwhile, GSTARX-SVR is the best model for forecasting inflow at one location of Bank Indonesia at Wes Java region. In general, these results in accordance with the third M3 forecasting competition conclusion, i.e. the more complicated model do not necessary yield better forecast than the simpler one.

Keywords: GSTARX, VARX, SVR, Inflow, Outflow

I. INTRODUCTION

The results of recently study of time series analysis showed that some data are not only related to previous events but also related to other neighbouring locations, known as have spatio-temporal relationship. One of time series models that involves the spatio-temporal aspect is the Generalized Space Time Autoregressive or GSTAR. GSTAR is an extension of Space-Time Autoregressive or STAR model. The advantage of the GSTAR model is the more flexible parameters than STAR model in explaining the interrelationships of different locations and times in the space time data.

It implies GSTAR can be applied to spatio-temporal data with heterogeneous location [1]. In practice, GSTAR model was applied in the field of economy, tourism and rainfall. In general, the methods in time series analysis are divided into two approaches, i.e. linear and nonlinear models. Previous study showed that inflow and outflow data in all Bank Indonesia Representative Offices contained both linear and non-linear components [2]. The hybrid model that combining linear and nonlinear models tended to improve the accuracy of the forecast [3]. Prediction of time series data with non-linear approach can be done by using Support Vector Machine (SVM). The concept of SVM is to find the best hyperplane that serves as a separator of two classes in input space. SVM is developed based on statistical learning theory and Structural Risk Minimization. SVM has demonstrated performance as a method that can overcome overfitting by minimizing the upper limit of generalization error, which is a powerful tool for supervised learning [4]. One of the SVM modifications that frequently be used for regression approaches is Support Vector Regression (SVR). Some researchers have applied SVR in many areas, including the economic field to predict stock price and return.

Inflow and outflow data in Bank Indonesia have great diversity on the scale of space and time. So, it requires a modeling method that considers location factors. In addition, due to nonlinear pattern in the spatio-temporal data especially inflow and outflow, the hybrid VARX-SVR and GSTARX-SVR modeling method is proposed and used for forecasting it. The VARX-SVR method is the combination of significant lag variables in VARX model as predictors in SVR model. Whereas, the GSTARX-SVR method is the combination of significant lag variables in GSTARX model as predictors in SVR model.

Hence, this study applied the VARX,GSTARX, VARX-SVR, and GSTARX-SVR methods for forecasting spatio-temporal data with exogenous variables. Both simulation and real data are used as case study. Finally, the forecast accuracy between these methods is compared and evaluated by using the root mean square of error (RMSE) and symmetric mean absolute percentage error (sMAPE).

II. METHODOLOGY

Data Sources and Research Variables

Two types of data are used in this study, i.e. (1) simulation data that be designed consist of trend, seasonal,

Revised Manuscript Received on February 05, 2019.

Suhartono, Department of Statistics, Faculty of Mathematics, Computing and Data Science, InstitutTeknologiSepuluhNopember, Kampus ITS Sukolilo Surabaya 60111, Indonesia,

Bahagiati Maghfiroh, Department of Statistics, Faculty of Mathematics, Computing and Data Science, Institut Teknologi SepuluhNopember, Kampus ITS Sukolilo Surabaya 60111, Indonesia

Santi Puteri Rahayu, Department of Statistics, Faculty of Mathematics, Computing and Data Science, InstitutTeknologiSepuluhNopember, Kampus ITS Sukolilo Surabaya 60111, Indonesia

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

and calendar variation effect, and (2) real data about inflow and outflow in Bank Indonesia at three locations at West Java, namely KBI Bandung, KBI Tasikmalaya, and KBI Cirebon. The inflow and outflow data are observed from January 2003 to December 2014. These data are divided into two parts, i.e. January 2003 to December 2013 for training dataset and period January to December 2014 for testing dataset.

Analysis Step

As aforementioned previously, the modeling of VARX, GSTARX, VARX-SVR, and GSTARX-SVR will be done through two data types, i.e. (1) simulation data, and (2) real data. In general, the steps of analysis in the simulation study are as follows:

Step 1: Determine the value of variance-covariance matrix $\mathbf{\Omega}$ where residuals are all correlated between locations, with different variance or $\sigma^{(ii)} \neq \sigma^{(jj)}$. The value of the variance-covariance matrix is given by

$$\mathbf{\Omega} = \begin{bmatrix} 0.75 & 0.20 & 0.35 \\ 0.20 & 0.40 & 0.25 \\ 0.35 & 0.25 & 0.45 \end{bmatrix}.$$

Step 2: Establish the vector model AR (1), with parameter coefficients at 3 locations are as follows:

$$\mathbf{\Phi_1} = \begin{bmatrix} 0.35 & 0.30 & 0.30 \\ 0.25 & 0.45 & 0.25 \\ 0.20 & 0.20 & 0.40 \end{bmatrix}.$$

Step 3: Determine two forms of simulated study scenarios, i.e.

a. Residuals that follow linear model, i.e. $N_t^{(1)} = 0.35N_{t-1}^{(1)} + 0.30N_{t-1}^{(2)} + 0.30N_{t-1}^{(3)} + a_t^{(1)}$, $N_t^{(2)} = 0.25N_{t-1}^{(1)} + 0.45N_{t-1}^{(2)} + 0.25N_{t-1}^{(3)} + a_t^{(2)}$, and $N_t^{(3)} = 0.20N_{t-1}^{(1)} + 0.20N_{t-1}^{(2)} + 0.40N_{t-1}^{(3)} + a_t^{(3)}$.

b. Residuals that follow nonlinear model, i.e.

$$N_t^{(1)} = 4N_{t-1}^{(1)} \times \exp\left\{-0.25[N_{t-1}^{(1)}]^2\right\} + 1.25N_{t-1}^{(2)} \times \exp\left\{-0.25[N_{t-1}^{(2)}]^2\right\} + 1.5N_{t-1}^{(3)} \times \exp\left\{-0.25[N_{t-1}^{(3)}]^2\right\} + a_t^{(1)}$$

$$N_t^{(2)} = 1.25N_{t-1}^{(1)} \times \exp\left\{-0.25[N_{t-1}^{(1)}]^2\right\} + 3.5N_{t-1}^{(2)} \times \exp\left\{-0.25[N_{t-1}^{(2)}]^2\right\} + 1.6N_{t-1}^{(3)} \times \exp\left\{-0.25[N_{t-1}^{(3)}]^2\right\} + a_t^{(2)}$$

$$N_t^{(3)} = 1.6N_{t-1}^{(1)} \times \exp\left\{-0.25[N_{t-1}^{(1)}]^2\right\} + 1.1N_{t-1}^{(2)} \times \exp\left\{-0.25[N_{t-1}^{(2)}]^2\right\} + 3N_{t-1}^{(3)} \times \exp\left\{-0.25[N_{t-1}^{(3)}]^2\right\} + a_t^{(3)}.$$

Step 4: Identify the dummy variable.

a. The trend component is a linear component, i.e. $T_t^{(i)} = \alpha_1^{(i)} T$. The coefficients for trends used in all scenarios are the same, i.e. $\alpha_1^{(1)} = 0.21$, $\alpha_1^{(2)} = 0.23$, $\alpha_1^{(3)} = 0.24$.

b. The seasonal component is obtained by the following equation,

$$S_t^{(i)} = \gamma_1^{(i)} S_{1,t} + \gamma_2^{(i)} S_{2,t} + \ldots + \gamma_{12}^{(i)} S_{12,t}.$$

where, $S_t^{(1)} = 22S_{1,t} + 24S_{2,t} + 27S_{3,t} + 24S_{4,t} + 22S_{5,t} + 17S_{6,t} + 13S_{7,t} + 10S_{8,t} + 6S_{9,t} + 10S_{10,t} + 13S_{11,t} + 17S_{12,t}$.

$$S_t^{(2)} = 25S_{1,t} + 28S_{2,t} + 30S_{3,t} + 28S_{4,t} + 25S_{5,t} + 21S_{6,t} + 16S_{7,t} + 13S_{8,t} + 9S_{9,t} + 13S_{10,t} + 16S_{11,t} + 21S_{12,t}.$$

$$S_t^{(3)} = 17S_{1,t} + 21S_{2,t} + 24S_{3,t} + 21S_{4,t} + 17S_{5,t} + 12S_{6,t} + 9S_{7,t} + 6S_{8,t} + 4S_{9,t} + 6S_{10,t} + 9S_{11,t} + 12S_{12,t}.$$

c. The component for calendar variation in this simulation study uses only the outflow data simulation model, so that the effect of variation of calendar used is on the occurrence of Eid and one month before the occurrence of Eid, with the following equation,

$$M_t^{(i)} = \delta_1^{(i)} M_{1,t} + \cdots + \delta_4^{(i)} M_{4,t} + \beta_1^{(i)} M_{1,t-1} + \cdots + \beta_4^{(i)} M_{4,t-1}.$$

where: $M_t^{(1)} = 32M_{1,t} + 43M_{2,t} + 47M_{3,t} + 49M_{4,t} + 50M_{1,t-1} + 46M_{2,t-1} + 39M_{3,t-1} + 34M_{4,t-1}$.

$$M_t^{(2)} = 45M_{1,t} + 49M_{2,t} + 55M_{3,t} + 57M_{4,t} + 60M_{1,t-1} + 54M_{2,t-1} + 44M_{3,t-1} + 36M_{4,t-1}.$$

$$M_t^{(3)} = 37M_{1,t} + 40M_{2,t} + 47M_{3,t} + 51M_{4,t} + 49M_{1,t-1} + 47M_{2,t-1} + 41M_{3,t-1} + 37M_{4,t-1}.$$

Step 5: Perform parameter estimation

a. The first level model is $\mathbf{Y}_t^{(i)} = \mathbf{T}_t^{(i)} + \mathbf{S}_t^{(i)} + \mathbf{M}_t^{(i)} + \mathbf{N}_t^{(i)}$

b. The second level model for VAR is $\mathbf{N}_t^{(i)} = \mathbf{\Phi}_{10}\mathbf{N}_{t-1}^{(i)} + \mathbf{\Phi}_{11}\mathbf{N}_{t-1}^{(i)} + \mathbf{a}_{t-1}^{(i)}$. And, the second level for GSTAR is $\mathbf{N}_t^{(i)} = \mathbf{\Phi}_{10}\mathbf{N}_{t-1}^{(i)} + \mathbf{\Phi}_{11}\mathbf{W}^{(1)}\mathbf{N}_{t-1}^{(i)} + \mathbf{a}_{t-1}^{(i)}$.

Step 6: Perform nonlinear model by applying SVR. Then compare the model accuracy of the four methods, i.e. VARX, VARX-SVR, GSTARX, and GSTARX-SVR.

The steps of analysis in the real data applications are starting from the identification step, include the determination of the effects of variation calendar through the time series plot of data, stationary examination, and the determination of AR order. Then, the process continues to the estimation parameter step, model testing, diagnostic check, and the forecasting.

III. RESULTS

Simulation Study

The simulation is done in two scenarios, namely linear residual and nonlinear residual. The time series plot of linear residual simulation data with trend, seasonal and calendar variation effects is shown in Figure 1.

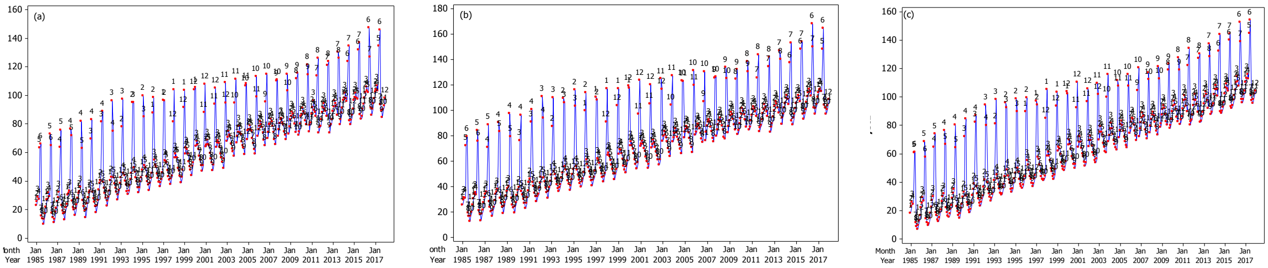


Fig. 1 Time series plot of simulation data with linear residual in (a) location 1, (b) location 2 and (c) location 3

Step one is to apply the first level modeling, i.e. fit time series regression to each data and the model for location one is

$$Y_t^{(1)*} = 0.21T_t^{(1)} + 22.02S_{1,t}^{(1)} + 23.97S_{2,t}^{(1)} + 27.01S_{3,t}^{(1)} + 24.12S_{4,t}^{(1)} + 22.1S_{5,t}^{(1)} + 17.18S_{6,t}^{(1)} + 13.52S_{7,t}^{(1)} + 10.26S_{8,t}^{(1)} + 6.24S_{9,t}^{(1)} + 10.03S_{10,t}^{(1)} + 13.20S_{11,t}^{(1)} + 17.13S_{12,t}^{(1)} + 32.41M_{1,t}^{(1)} + 43.57M_{2,t}^{(1)} + 46.44M_{3,t}^{(1)} + 48.61M_{4,t}^{(1)} + 49.91M_{1,t-1}^{(1)} + 46.39M_{2,t-1}^{(1)} + 38.72M_{3,t-1}^{(1)} + 32.34M_{4,t-1}^{(1)} + N_t^{(1)} .$$

where $N_t^{(1)}$ is error component of the first level model. By using the same approach, the calculation of $N_t^{(i)}$ for i=1,2,3 are obtained. Then, the second level is modeling the error component by applying VAR and GSTAR models. The estimation result of VAR(1) model for all three locations are as follows:

$$\begin{bmatrix} N_t^{(1)} \\ N_t^{(2)} \\ N_t^{(3)} \end{bmatrix} = \begin{bmatrix} 0.278 & 0.286 & 0.378 \\ 0.119 & 0.421 & 0.443 \\ 0.134 & 0.197 & 0.480 \end{bmatrix} \begin{bmatrix} N_{t-1}^{(1)} \\ N_{t-1}^{(2)} \\ N_{t-1}^{(3)} \end{bmatrix} + \begin{bmatrix} a_t^{(1)} \\ a_t^{(2)} \\ a_t^{(3)} \end{bmatrix}.$$

Then, GSTAR modeling is performed with the same order of VAR. In this research, two estimation methods of GSTAR are applied and compared, i.e. OLS and GLS. The results showed that the GLS method yielded is more efficient estimators than OLS and it is in line with the results of some previous studies [5, 6]. Hence, the results of GSTAR (1_1)-GLS model is as follows:

$$\begin{bmatrix} N_t^{(1)} \\ N_t^{(2)} \\ N_t^{(3)} \end{bmatrix} = \begin{bmatrix} 0.357 & 0.285 & 0.285 \\ 0.249 & 0.452 & 0.249 \\ 0.211 & 0.211 & 0.370 \end{bmatrix} \begin{bmatrix} N_{t-1}^{(1)} \\ N_{t-1}^{(2)} \\ N_{t-1}^{(3)} \end{bmatrix} + \begin{bmatrix} a_t^{(1)} \\ a_t^{(2)} \\ a_t^{(3)} \end{bmatrix}.$$

Finally, nonlinear modeling is done by applying VARX-SVR and GSTARX-SVR. Figure 2 are the matrix plot of residual models atthe first step from both simulation scenarios, i.e. linear correlated and nonlinear correlate residuals, respectively.

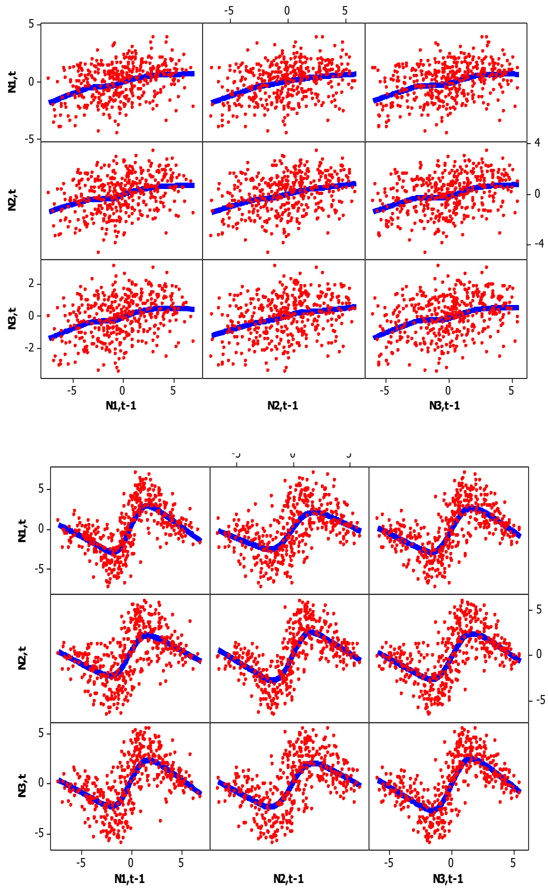


Fig. 2 Matrix plot of linear correlated residual (left) and nonlinear correlated residual (right)

The matrix plot in Figure 3 shows that the residual from the simulation study already satisfied the original patterns. In this study, VARX-SVR and GSTARX-SVR were applied to both linear and nonlinear correlated residuals to know whether these methods could improve the forecast accuracy. The grid search method was employed to get the optimum parameter estimation for each model. Moreover, the result of the VARX-SVR model for location one is

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$\hat{Y}_t^{(1)} = \sum_{i=1}^{383}(\beta_i - \beta_i^*)K(x_t, \mathbf{x}) + b$$
$$= 350\exp\left(-\frac{1}{2(0.05)^2}\|x_1 - \mathbf{x}\|^2\right) - 289.55\exp\left(-\frac{1}{2(0.05)^2}\|x_2 - \mathbf{x}\|^2\right) + 350\exp\left(-\frac{1}{2(0.05)^2}\|x_3 - \mathbf{x}\|^2\right) + \cdots - 350\exp\left(-\frac{1}{2(0.05)^2}\|x_{381} - \mathbf{x}\|^2\right) +$$
$$350\exp\left(-\frac{1}{2(0.05)^2}\|x_{382} - \mathbf{x}\|^2\right) - 350\exp\left(-\frac{1}{2(0.05)^2}\|x_{383} - \mathbf{x}\|^2\right) - 0.59$$

Whereas, the GSTARX-SVR model for location one is

$$\hat{Y}_t^{(1)} = \sum_{i=1}^{383}(\beta_i - \beta_i^*)K(x_t, \mathbf{x}) + b$$
$$= 350,9\exp\left(-\frac{1}{2(0,05)^2}\|x_1 - \mathbf{x}\|^2\right) - 350,9\exp\left(-\frac{1}{2(0,05)^2}\|x_2 - \mathbf{x}\|^2\right) + 350,9\exp\left(-\frac{1}{2(0,05)^2}\|x_3 - \mathbf{x}\|^2\right) + \cdots + 350,9\exp\left(-\frac{1}{2(0,05)^2}\|x_{381} - \mathbf{x}\|^2\right) +$$
$$350,9\exp\left(-\frac{1}{2(0,05)^2}\|x_{382} - \mathbf{x}\|^2\right) + 350,9\exp\left(-\frac{1}{2(0,05)^2}\|x_{383} - \mathbf{x}\|^2\right) - 0,59$$

After fitting all four models, then determination of the best model is done by comparing RMSE at testing datasets. Figure 3 shows the boxplot of the RMSE models at testing dataset in both linear and nonlinear simulation scenarios. The red, green, blue, and orange color represents the result of VARX, VARX-SVR, GSTARX, and GSTARX-SVR model, respectively. The results show that VARX-SVR model has the smallest RMSE in the linear residual simulation scenario. Moreover, the results in nonlinear residual simulation scenario show that GSTARX-SVR and VARX-SVR yield similar results and better than VARX and GSTARX models. These results indicate that both VARX-SVR and GSTARX-SVR models are appropriate for forecasting spatio-temporal data that have both linear and nonlinear correlated patterns.

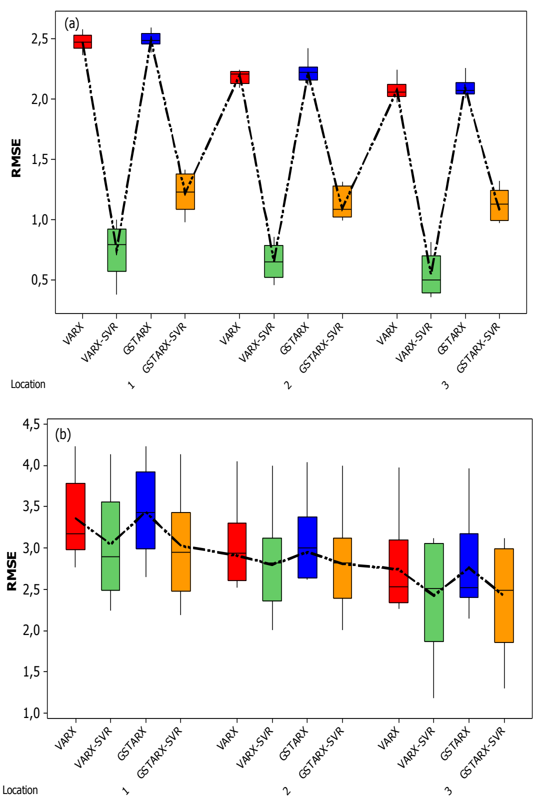


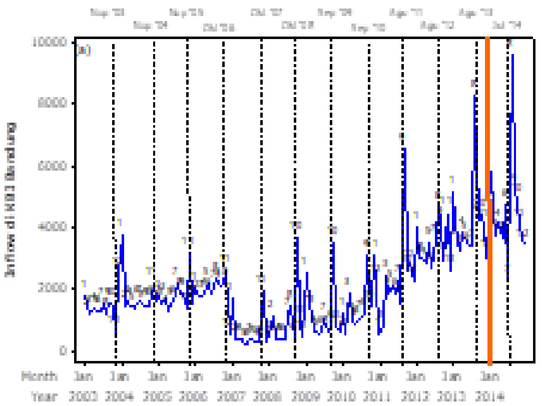
Fig. 3 The RMSE of both linear residual (left) and non-linear residuals (right) models at testing data

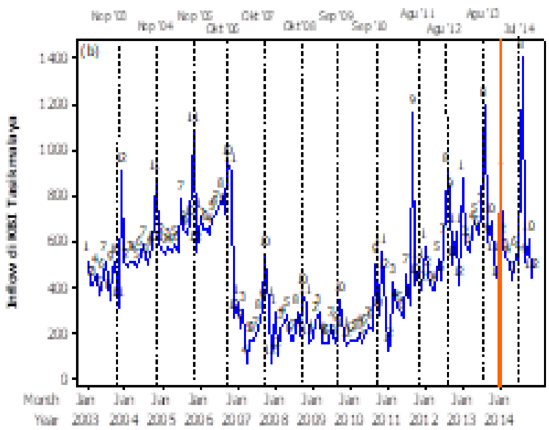
In general, the results of simulation study show that the VARX-SVR and GSTARX-SVR models as the more complex models tend to give more accurate forecast than the VARX and GSTARX models as the simpler models.

Moreover, these results also proved that VARX-SVR and GSTARX-SVR could handle accurately both linear and nonlinear pattern in the spatio-temporal data.

Real Case Study

The modeling of four models in real data, i.e. inflow and outflow, is conducted by following the same steps at the simulation study. Figure 4 and 5 show the time series plot of inflow and outflow, respectively, at three location of Bank Indonesia in West Java region. Moreover, there are three types of the spatial weight for implementing both GSTARX and GSTARX-SVR models that be used and compared to find the best model, i.e. uniform, inverse distance, and normalization inference of partial cross correlation (NIPCC) weights.





Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

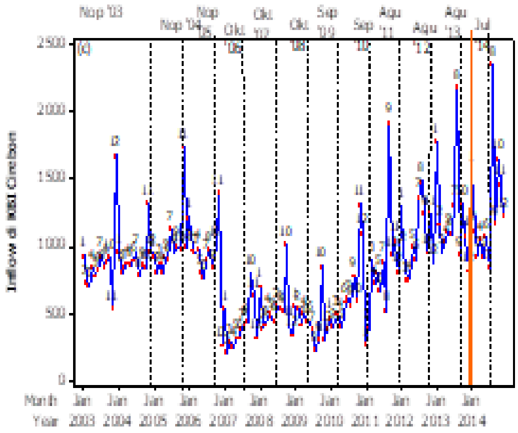


Fig. 4 Time series plot of inflow data at (a) KBI Bandung, (b) KBI Tasikmalaya, and (c) KBI Cirebon

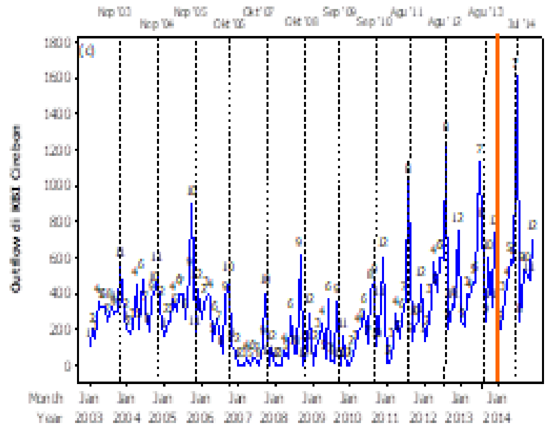
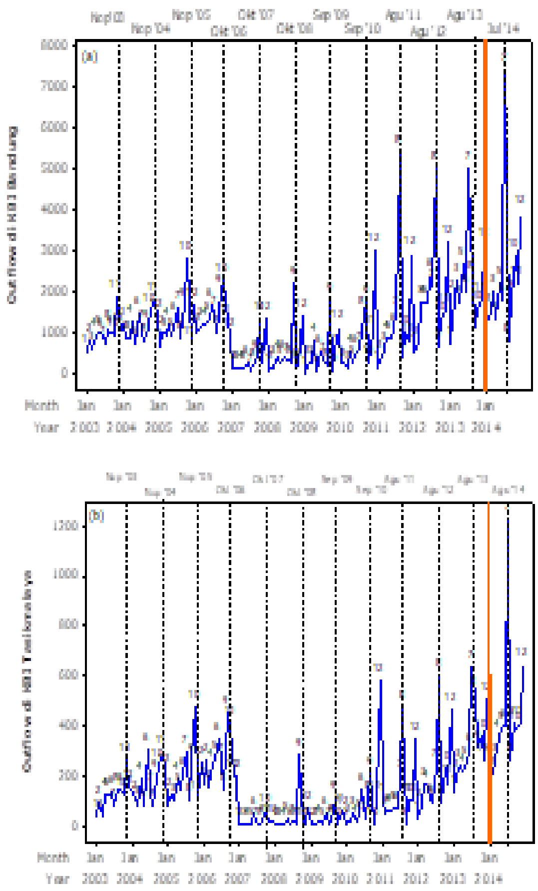


Fig. 5 Time series plot of outflow data at (a) KBI Bandung, (b) KBI Tasikmalaya, and (c) KBI Cirebon



The first step, inflow and outflow data are modeled by Time Series Regression and obtained the residuals that not satisfy white noise condition. It means that the modeling process could be continued at the second steps by applying for models, i.e. VARX, GSTARX, VARX-SVR, and GSTARX-SVR. Moreover, evaluation the pattern of residuals is done by employing neglected nonlinearity test and the results are presented at Table 1. These results indicate that the residuals of the first step model, both for inflow and outflow, following nonlinear relationship.

Table. 1 The results of Neglected Nonlinearity Test for Residuals at the First Step

Data	Location	df	χ^2	p-value
Outflow	KBI Bandung	2	31.2081	0.0000
	KBI Tasikmalaya	2	63.3463	0.0000
	KBI Cirebon	2	33.5406	0.0000
Inflow	KBI Bandung	2	61.0345	0.0000
	KBI Tasikmalaya	2	40.9906	0.0000
	KBI Cirebon	2	56.0695	0.0000

Furthermore, summary of the results is illustrated only for outflow data at KBI Bandung and for GSTARX model only for uniform weight. The model for the first level by using Time Series Regression for outflow data at KBI Bandung is

$$Y_t^{(1)*} = 19.68T_t^{(1)} - 1394.99D_{1,t}^{(1)} - 3263.84D_{2,t}^{(1)} + 20.04TD_{2,t}^{(1)} + 295.21S_{3,t}^{(1)} + 351.61S_{4,t}^{(1)} + 315.21S_{5,t}^{(1)} + 609.57S_{6,t}^{(1)} + 687.52S_{7,t}^{(1)} + 314.56S_{8,t}^{(1)} + 1239.59S_{12,t}^{(1)} + 1104.46M_{2,t}^{(1)} + 2490.87M_{3,t}^{(1)} + 2404.38M_{4,t}^{(1)} + 2205.91M_{1,t-1}^{(1)} + 1252.35M_{2,t-1}^{(1)} + N_t^{(1)}.$$

From the residual of the first level, the result of linear modeling by using VAR(1) at KBI Bandung is

$$N_t^{(1)*} = 0.1397N_{t-1}^{(1)} - 1.3793N_{t-1}^{(2)} - 0.0972N_{t-1}^{(3)} + a_t^{(1)}$$

and the corresponding result for VARX-SVR is

$$\hat{Y}_t^{(1)} = \sum_{t=1}^{132}(\beta_t - \beta_t^*)K(\mathbf{x}_t,\mathbf{x}) + b$$

$$= 351.5\exp\left(-\frac{1}{2(0.05)^2}\|\mathbf{x}_1-\mathbf{x}\|^2\right) + 193.12\exp\left(-\frac{1}{2(0.05)^2}\|\mathbf{x}_2-\mathbf{x}\|^2\right) + 351.5\exp\left(-\frac{1}{2(0.05)^2}\|\mathbf{x}_3-\mathbf{x}\|^2\right) + \cdots - 201.92\exp\left(-\frac{1}{2(0.05)^2}\|\mathbf{x}_{130}-\mathbf{x}\|^2\right) +$$

$$-110.51\exp\left(-\frac{1}{2(0.05)^2}\|\mathbf{x}_{131}-\mathbf{x}\|^2\right) - 351.5\exp\left(-\frac{1}{2(0.05)^2}\|\mathbf{x}_{132}-\mathbf{x}\|^2\right) - 34.081.$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Whereas, the result of GSTAR (1_1) model with uniform weight is

$$N_t^{(1)*} = -0.067N_{t-1}^{(1)} + 0.039N_{t-1}^{(2)} + 0.039N_{t-1}^{(3)} + a_t^{(1)}$$

and the corresponding result for GSTARX-SVR is

$$\hat{Y}_t^{(1)} = \sum_{i=1}^{132}\left(\beta_i - \beta_i^*\right)K\left(x_t, \mathbf{x}\right) + b$$

$$= 354.3\exp\left(-\frac{1}{2(0.05)^2}\left\|x_1 - \mathbf{x}\right\|^2\right) + 190.51\exp\left(-\frac{1}{2(0.05)^2}\left\|x_2 - \mathbf{x}\right\|^2\right) + 354.3\exp\left(-\frac{1}{2(0.05)^2}\left\|x_3 - \mathbf{x}\right\|^2\right) + \cdots - 204.54\exp\left(-\frac{1}{2(0.05)^2}\left\|x_{130} - \mathbf{x}\right\|^2\right) +$$

$$-113.12\exp\left(-\frac{1}{2(0.05)^2}\left\|x_{131} - \mathbf{x}\right\|^2\right) - 354.3\exp\left(-\frac{1}{2(0.05)^2}\left\|x_{132} - \mathbf{x}\right\|^2\right) - 36,69.$$

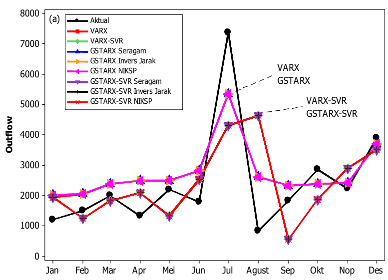
Outflow at KBI Bandung is affected by Eid-Fitrevent as calendar variation at two period, i.e. on the month of Eid-Fitr at all weeks (1,2,3,4)an on 1 month before Eid-Fitr when the date of this event is at week 1 and 2. Based on the models formed for outflow at each location, it is known that there is a spatial effect in the data, i.e. outflow at certain location is influenced by outflow at other locations. It means that outflow at KBI Bandung is influenced by outflow of KBI Tasikmalaya and Cirebon one month earlier. Similarly, outflow at KBI Tasikmalaya and Cirebon have the same model interpretation as at KBI Bandung.

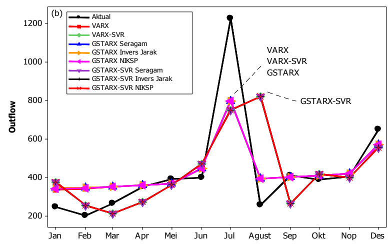
By using the same steps, the models for inflow and outflow at three KBI in West Java region are obtained and the summary of accuracy model based on sMAPE are shown at Table 2. The results at Table 2 show that the best model for forecasting outflow at three locations of KBI is VARX model. Moreover, VARX also yields more accurate forecast than other methods for forecasting inflow at two KBI, i.e.Bandung and Cirebon. Otherwise, the GSTARX-SVR model as one of proposed hybrid method give most accurate forecast at KBI Tasikmalaya.

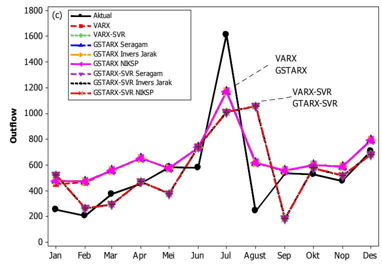
Table. 2 The results of Four models for forecasting inflow and outflow at three locations

Data	Method	Weight	sMAPE Testing		
			Bandung	Tasikmalaya	Cirebon
Outflow	VARX*		0.335917	0.195800	0.332762
	VARX-SVR		0.479572	0.295180	0.397689
	GSTARX	Uniform	0.339864	0.200129	0.338583
		Inverse Distance	0.339804	0.200066	0.338522
		NIPCC	0.338752	0.197526	0.337853
	GSTARX-SVR	Uniform	0.479572	0.295180	0.397689
		Inverse Distance	0.479572	0.295180	0.397689
		NIPCC	0.479572	0.295180	0.397689
Inflow	VARX		0.225814	0.383331	0.212051
	VARX-SVR		0.295873	0.263470	0.220199
	GSTARX	Uniform	0.226018	0.383364	0.212077
		Inverse Distance	0.226050	0.383370	0.212094
	GSTARX-SVR	Uniform	0.295832	0.263445	0.220199
		Inverse Distance	0.295816	0.263450	0.220199

In general, the results show that VARX-SVR and GSTARX-SVR as non-linear and complex model do not give better results than VARX and GSTARX as linear and simpler model. These results support to one of the M3 competition results, i.e. the more complicated method is not necessarily better than the simple method [7]. In addition, these predictions evaluation are used*k*-step-ahead forecast so that sMAPE of non-linear and complex model tend to worse at testing dataset due to overfitting as shown at Figure 6.







Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

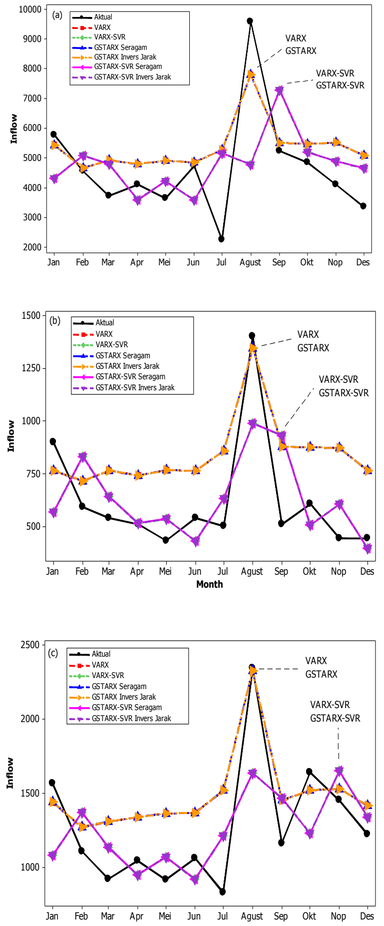


Fig. 6 The *k*-step-ahead forecast of outflow (top) and inflow (down) with non-linear model

Moreover, the results at Figure 6 show that both VARX and GSTARX tend to produce similar forecast for outflow and inflow data. It indicates the forecast lines of both models are coincide. Otherwise, the results of both VARX-SVR and GSTARX-SVR as non-linear and complex models are also coincide and tend to less accurate than the forecast of both VARX and GSTARX models.

IV. CONCLUSIONS

Based on the results at the aforementioned above, it could be concluded that the proposed hybrid VARX-SVR and GSTARX-SVR as nonlinear spatio-temporal model could work well for forecasting both linear and nonlinear pattern data. It was shown by the results of simulation study that these both proposed models yielded more accurate forecast than VARX and GSTARX as linear classical models. These results in line with the M3 and M4 forecasting competition results, i.e. hybrid method that combining some methods tend to give better forecast than individual methods [8].

Moreover, the results at real data showed that VARX model as a simple model yielded more accurate forecast than VARX-SVR and GSTARX-SVR as a more complex model in forecasting inflow and outflow data in Indonesia. Again, this result also concordance with the M3 forecasting competition results, i.e. complex methods do not necessary give better result than the simpler one. These results were based on *k*-step-ahead forecast and usually give not accurate prediction for long period. Further research is needed to validate the accuracy of the VARX-SVR and GSTARX-SVR for short period forecast, particularly for solving now casting problem that many researchers did in recent years.

V. ACKNOWLEDGEMENT

This research was funded by a LPPM-ITS under scheme of “PenelitianPascasarjana”, project No. 1337/PKS/ITS/ 2018. The authors thank to the General Director of DIKTI for funding and to anonymous referees for their useful suggestions.

REFERENCES

1. Suhartono, Prastyo, D.D., Kuswanto, H., & Lee, M.H. (2018). Comparison between VAR, GSTAR, FFNN-VAR and FFNN-GSTAR Models for Forecasting Oil Production. *MATEMATIKA*. Vol. 34(1), pp. 103-11.
2. Prayoga, I.G.S.A., Suhartono, Rahayu, S.P. (2017). Top-Down Forecasting for High Dimensional Currency Circulation Data of Bank Indonesia. *International Journal of Advances in Soft Computing and its Applications*. Vol. 9(2), pp. 62-74.
3. Zhang, G. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*. Vol.50, pp. 159-175.
4. Guo, J., Yi, P., Wang, R., Ye, Q., &Zhao, C.(2014). Feature Selection for Least Square Projection Twin Support Vector Machine. *Neurocomputing*. Vol. 14, pp. 174-183.
5. Suhartono,Wahyuningrum, S.R., Setiawan, Akbar, M.S. (2016). GSTARX-GLS Model for Spatio-Temporal Data Forecasting. *Malaysian Journal of Mathematical Sciences*, Vol. 10(S), pp. 91-103.
6. Setiawan,Suhartono,&Prastuti, M. (2016). S-GSTAR-SUR model for seasonal spatio temporal data forecasting. *Malaysian Journal of Mathematical Sciences*. Vol. 10(S), pp.53-65.
7. Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting*. Vol. 16(4), pp. 451-476.
8. Makridakis, S., Spiliotis, E.&Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*. [Forthcoming]