GSTARIX Model for Forecasting Spatio-Temporal Data with Trend, Seasonal and Intervention

M A Novianto¹, Suhartono², D D Prastyo³, A Suharsono⁴, and Setiawan⁵

¹,²,³,⁴,⁵ Department of Statistics, Faculty of Mathematics, Computing, and Data Science, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

suhartono@statistika.its.ac.id

Abstract. Generalized Space-Time Autoregressive (GSTAR) is a statistics model that is usually applied for forecasting data that have both spatial and temporal dependency. The monthly tourist arrival data in some locations are an example of spatio-temporal data. Most of the previous research on GSTAR model only focused on stationary data. Otherwise, tourist arrival data in Indonesia mostly contain trend, seasonal, and some extreme values caused by interventions or outliers. The objective of this study is to apply and develop GSTAR model for forecasting spatio-temporal data with trend, seasonal, and interventions or outliers. This model is then known as GSTAR with exogeneous variables or GSTARIX model. Then, the forecast accuracy of GSTARIX model is compared to VAR with exogenous variables or VARIX model. Monthly data about number of tourist arrivals to Jakarta, Surakarta, Surabaya, and Denpasar are used as case study. Moreover, two methods are used for parameter estimation, i.e. Ordinary Least Square (OLS) and Generalized Least Square (GLS). The criteria for selecting the best model is Root Mean Square Error (RMSE). The results showed that the best model for forecasting tourist arrivals in each location are different. The best model for forecasting number of tourist arrivals to Jakarta and Surabaya is GSTARIX-OLS. Whereas, the best model for Denpasar and Surakarta data are VARIX and GSTARIX-GLS, respectively.

1. Introduction

Space Time Autoregressive (STAR) model is a time series model developed in several locations simultaneously that accommodates the interconnection of space and time [1]. However, the STAR model is only suitable for locations with homogeneous characteristics, since this model assumes the same value parameters for all locations. Then Borovkova et al. introduced the Generalized Space Time Autoregressive or GSTAR model to address the heterogeneous inter-location characteristics [2]. Ruchjana et al. also used the GSTAR model for forecasting oil production [3]. Suhartono and Atok conducted a comparison of VARIMA and GSTAR models [4]. Subsequently, Suhartono and Subanar examined the optimal weight selection on the GSTAR model [5].

Some researchers incorporate seasonal elements into the GSTAR model. Wutsqa and Suhartonousing Seasonal VAR-GSTAR model for tourism data forecasting [6], and also Setiawan et al. developed Seasonal GSTAR model for forecasting the data of foreign tourist arrivals [7]. Other researchers used the GSTAR model for non-stationary data such as Nisak et al. used the GSTARIMA model for rainfall forecasting [8]. Messakh et al. also used the GSTARIMA model for forecasting rice production [9], and Bonar et al. using the GSTARI-ARCH model for Consumer Price Index [10]. The
addition of exogenous variables into the GSTAR model also made some researchers to increase the accuracy of forecasting. Ditago and Suhartono simulated the GSTAR-X model with calendar variations effect [11], Mubarak and Suhartono simulated GSTAR-X with metric predictor known as transfer function [12], Suhartono et al. forecasted inflation in East Java with non metric predictor known as intervention variable[13].

Tourist arrival data has unique characteristics because it contains a combination of trend, seasonal and intervention influences. The trend is a long-term component that reflects the growth or decrease of time series data over the data period. The trend results in nonstationary data at the level or should be done differencing so that it appears integrated element on GSTAR model. Seasonal is a periodic and repetitive pattern caused by several factors such as climate, holiday, and promotion. Intervention is the emergence of the effects of an event, both internal and external those are expected to affect the movement of time series data. Internal factors are factors that can be controlled, for example, government policy. While, the external factor is something that can not be controlled such as bomb incidents, forest fires, and others. The intervention variables used in this study are some events such as monetary crisis and bomb incidents in Indonesia combined with outlier detection.

This research proposes GSTAR model for forecasting spatio-temporal data containing trend, seasonal and interventions analysis for handling outliers which never done in previous research. This model then is known as GSTARIX model. To find out the forecast accuracy, the GSTARIX model is compared to the VARIX model by using Root Mean Square of Error (RMSE) criterion. The data used in this study are the number of foreign tourist arrival through four arrival gates in Java and Bali, i.e. Jakarta, Surakarta, Surabaya, and Denpasar. The data are observed for 25 years period, from 1996 to 2016, and divided into two parts, i.e. in-sample data from 1996 to 2015 and out-sample data for 2016 observations.

2. Literature review
In this section, the GSTAR and VAR that involve exogeneous variables, known as GSTARX and VARX, respectively, are presented. These models consist of two-level modelling, i.e. the first is modelling of trend, seasonal, and interventions or outliers using Time Series Regression or TSR method, and the second is modelling the residual of TSR by employing GSTAR and VAR models. The forecast of the first and second steps are combined to obtain the forecast of GSTARX and VARX models.

2.1. Autoregressive Integrated Moving Average (ARIMA) Model
In this study, ARIMA model is performed for detecting outliers in time series that usually imply the residuals could not fulfill normal distribution assumption. The ARIMA model can be written as [14]:

\[ \phi_p(B)(1 - B)^d Z(t) = \theta_q(B)a(t) \]  

where \( \phi_p(B) \) is AR operator, \( \theta_q(B) \) is MA operator, and \( a(t) \) is white noise residual with mean zero and variance \( \sigma^2_a \).

2.2. Time Series Regression Model
Time Series Regression or TSR model used in this study is the multiple linear regression where the predictors are dummy variables of trends, seasonal, and interventions or outliers. Moreover, the results of outlier detection based on ARIMA model are used as one of dummy variables in this TSR model. In general, TSR model with predictors are dummy variables of trend, seasonal, and outliers, can be written as

\[ Z_i = \alpha T + \sum_{m=1}^{d} \beta_m S_{m,i} + \sum_{j=1}^{g_i} \gamma_j I_{j,i} + N_i \]  

where \( \alpha \) is linear trend parameter, \( T \) is dummy variable for trend, \( \beta \) is seasonal parameter, \( S_{m,i} \) is dummy variable for seasonal pattern, \( \gamma \) is outlier parameter, \( I_{j,i} \) is dummy variable for outlier, and \( N_i \) is residual. The residual of TSR model usually has autocorrelation even though it is free from the effects of trend, seasonal and outlier.
2.3. Vector Autoregressive Model

In this study, Vector Autoregressive or VAR model is used for modelling the residual of TSR model that has autocorrelation. The VAR is a multivariate time series model that can be written as [14]:

\[ \Phi(B)D(B)\mathbf{N}(t) = \mathbf{e}(t) \]  

(3)

where \( \mathbf{N}(t) \) is the residual of TSR model. The building of VAR model is done through several steps such as the Box-Jenkins procedure, i.e. identification, parameter estimation, diagnostic checks, and forecasting [4]. Time series plot, Matrix Cross Correlation Function or MCCF, Matrix Partial Cross-Correlation Function and MPCCF, and Akaike’s Information Criterion or AIC value are used in identification step for selecting the order or the model. Least square method is applied at parameter estimation step. Then, diagnostic checks are employed to test whether the residuals of the model has satisfied white noise assumption. Finally, the calculation of the final prediction is done by the best model.

2.4. Generalized Space Time Autoregressive Model

As VAR model, GSTAR model in this study is used for modelling the residual of the TSR model, \( \mathbf{N}(t) \), that usually have autocorrelation. GSTAR is a special form of VAR model that can show separately the influence of time and spatial parameters. The GSTAR model for \( n \) locations, time order is \( p \) and spatial order are \( \lambda_1, \lambda_2, \ldots, \lambda_s \), or known as GSTAR \((p, \lambda_1, \lambda_2, \ldots, \lambda_s)\) can be written as [2]:

\[ \mathbf{N}(t) = \sum_{k=1}^{p} \Phi^k + \sum_{l=1}^{s} \Theta^l \mathbf{W}^{(l)}\mathbf{N}(t-k) + \epsilon(t) \]  

(4)

where \( \Phi^k = \text{diag}(\phi_{k0}, \phi_{k1}, \ldots, \phi_{kn}), \Theta^l = \text{diag}(\phi_{l0}, \phi_{l1}, \ldots, \phi_{ln}) \), \( k \) is order of temporal, \( l \) is order of spatial, \( \epsilon(t) \) is residual model that fulfills identically, independent, normally distributed with mean 0 and covariance \( \Sigma \). For example, GSTAR model with time and spatial order equal one can be written as:

\[ \mathbf{N}(t) = \left[ \Phi^t + \Theta^1 \mathbf{W}^{(1)} \right] \mathbf{N}(t-1) + \epsilon(t) \]  

(5)

If the residual of the TSR model is not stationary in mean then differencing process should be applied to this residual. This integrated element from differencing process is modelled by GSTAR and it is known as GSTARI model. In matrix notation, GSTARI model with time and spatial order equal one can be written as:

\[ \mathbf{N}'(t) = \left[ \Phi^t + \Theta^1 \mathbf{W}^{(1)} \right] \mathbf{N}'(t-1) + \epsilon(t) \]  

(6)

where \( \mathbf{N}'(t) = \mathbf{N}(t) - \mathbf{N}(t-1), \mathbf{N}'(t-1) = \mathbf{N}(t-1) - \mathbf{N}(t-2) \). For example, Equation (6) for three locations can be written as:

\[ \begin{bmatrix} N'_1(t) \\ N'_2(t) \\ N'_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \phi_{20} & 0 \\ 0 & 0 & \phi_{30} \end{bmatrix} + \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_{21} & 0 \\ 0 & 0 & \phi_{31} \end{bmatrix} \mathbf{W} \begin{bmatrix} w_{12} & w_{13} \\ w_{21} & w_{23} \\ w_{31} & w_{32} \end{bmatrix} \begin{bmatrix} N'_1(t-1) \\ N'_2(t-1) \\ N'_3(t-1) \end{bmatrix} + \begin{bmatrix} \epsilon_1(t) \\ \epsilon_2(t) \\ \epsilon_3(t) \end{bmatrix} \]  

(7)

Several matrices of spatial weights or \( \mathbf{W} \) could be used in GSTAR model, i.e. uniform, inverse of distance, normalization of cross-correlation (NCC), and normalization of inference of partial cross-correlation (NIPCC) weights [4].

2.5. Ordinary Least Square and Generalized Least Square Method

The methods that usually be used for parameter estimation in the GSTAR model are Ordinary Least Square (OLS) and Generalized Least Square (GLS). OLS is used when the residuals between equations or locations are uncorrelated, whereas GLS is used when the residuals between locations are correlated. According to Greene [15], GLS model with \( n \) equations or locations where each equation contains of \( k \) predictors can be written as follows:
\[ Z_i = \beta_{i0} + \beta_{i1}X_{i1} + \beta_{i2}X_{i2} + \ldots + \beta_{im}X_{im} + e_i \]

\[ Z_j = \beta_{j0} + \beta_{j1}X_{j1} + \beta_{j2}X_{j2} + \ldots + \beta_{jm}X_{jm} + e_j. \]  

The assumptions of GLS method are \( E(e) = 0 \) and \( E(e'e) = \sigma_q I_t \), where \( i, j = 1, 2, \ldots, n \).

### 2.6. Parameter Estimation of GSTAR Model

Let \( N'(t) \) is multivariate residual from TSR model of three locations, then GSTAR\([1,1]\)-I(1) model can be written as [7]:

\[
N'(t) = [\Phi^0 + \Phi^1 W(t)] N(t - 1) + e(t)
\]  

where \( N'(t) = N(t) - N(t - 1), N'(t - 1) = N(t - 1) - N(t - 2) \), or in matrix representation as:

\[
\begin{bmatrix}
N'_1(t) \\
N'_2(t) \\
N'_3(t)
\end{bmatrix} =
\begin{bmatrix}
\Phi^0 & 0 & 0 \\
0 & \Phi^0 & 0 \\
0 & 0 & \Phi^0
\end{bmatrix} \begin{bmatrix}
N'_1(t - 1) \\
N'_2(t - 1) \\
N'_3(t - 1)
\end{bmatrix} +
\begin{bmatrix}
\phi_{10} \\
\phi_{20} \\
\phi_{30}
\end{bmatrix} \begin{bmatrix}
0 \\
w_{12} \\
w_{13}
\end{bmatrix} +
\begin{bmatrix}
e_{12} \\
e_{22} \\
e_{32}
\end{bmatrix}.
\]

Hence, for each \( i = 1, 2, \ldots, n \), it implies that

\[ N'(t) = \Phi N'(t - 1) + \Phi' V(t - 1) + e(t). \]

The GSTARI-GLS model can be written as:

\[
Y_i(t) = X_i(t)\beta_i + e_i(t)
\]

where

\[
Y_i(t) = N'_i(t), X_i(t) = [N'_1(t - 1) \quad V'_1(t - 1)], \beta_i = [\phi_{i0} \quad \phi_{i1} \quad \phi_{i2} \quad \phi_{i3}], e_i(t) = e_i(t).
\]

Hence, the GSTARI-GLS model in matrix form could be written as follows:

\[
\begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_{n,n}
\end{bmatrix} =
\begin{bmatrix}
X_1 & 0 & \cdots & 0 \\
0 & X_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & X_{n,n}
\end{bmatrix} \begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_n
\end{bmatrix} +
\begin{bmatrix}
e_{11} \\
e_{12} \\
\vdots \\
e_{n,n}
\end{bmatrix}.
\]

The assumption of the residuals of GSTARI-GLS model are uncorrelated in each location, i.e.

\[ E(e_e,e_s) = \begin{bmatrix} 0 & \sigma_{e_0,t} \\
\sigma_{e_0,t} & 0 \\
\end{bmatrix} \]

where \( i, j = 1, 2, \ldots, n \) and \( t, s = 1, 2, \ldots, T \). Otherwise, the residuals of GSTARI-GLS have relationship between locations or equations. Thus, variance-covariance matrix of residual is

\[
E(e'e') = \sigma_q \Sigma = \begin{bmatrix}
\sigma_{11} & \cdots & \sigma_{1n} \\
\sigma_{21} & \cdots & \sigma_{2n} \\
\vdots & \ddots & \vdots \\
\sigma_{n1} & \cdots & \sigma_{nn}
\end{bmatrix} \begin{bmatrix}
\Sigma \\
\Sigma \\
\vdots \\
\Sigma
\end{bmatrix} = \Omega.
\]

\[ \Omega I_t = \Sigma \otimes I_t = \Omega. \]
where $\Omega$ is matrix $(n \times T) \times (n \times T)$. Parameter estimation in GSTARI-GLS is done by minimizing generalized sum of square error $\epsilon'\Omega^{-1}\epsilon$. The result of GLS estimation in GSTARI-GLS model is

$$\hat{\beta} = (X\Omega^{-1}X)^{-1}X\Omega^{-1}Y.$$ 

Since $\Omega = \Sigma \otimes I_T$, then $\hat{\beta}$ estimator is

$$\hat{\beta} = (X(\Sigma \otimes I_T)^{-1}X)^{-1}(X(\Sigma \otimes I_T)^{-1}Y).$$ (13)

2.7. VARIX and GSTARIX Model

VARIX and GSTARIX model is a combination of TSR at the first step and VARI and GSTARI at the second step, respectively. Forecast calculation of VARIX and GSTARIX models are

$$\hat{Z}_i(t) = \hat{Z}_i(t) + \hat{N}_i(t)$$

where $\hat{Z}_i(t)$ is forecast of VARI or GSTARIX model, $\hat{Z}_i(t)$ is forecast of TSR Model, and $\hat{N}_i(t)$ is forecast of VARI or GSTARI model.

3. Result and discussion

3.1. Autoregressive Integrated Moving Average (ARIMA) Model

Figure 1 shows about the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) of number of tourist arrivals data to four locations, i.e. Jakarta, Denpasar, Surabaya, and Surakarta, after implementing non-seasonal differencing or $d=1$. These graphs show that ACF the data still not stationary. It illustrates by the ACF plotstend to have significance values at seasonal lags. Thus, it needs seasonal differencing or $D=1$, $S=12$.

Furthermore, the ACF of differencing series both non-seasonal and seasonal differencing ($d=1$, $D=1$, $S=12$) mostly tend to follow cuts off pattern both in non-seasonal and seasonal lags, and the PACF mostly tend to have dies down pattern. Hence, MA order both in non-seasonal and seasonal components seem more appropriate for these series. The results of parameter estimation and significance test of ARIMA model in each location are shown at Table 1.

| Location  | ARIMA   | Parameter | Estimation | $p$-value | WhiteNoise | KS($p$-value) |
|-----------|---------|-----------|------------|-----------|------------|--------------|
| Jakarta   | (0,1,1) | $\theta_1$ | 0.628      | $<0.001$  | Yes        | 0.063        |
|           |         | $\theta_3$ | 0.607      | $<0.001$  |            | (0.03)       |
| Denpasar  | (0,1,4) | $\theta_1$ | 0.543      | $<0.001$  | Yes        | 0.09         |
|           |         | $\theta_4$ | 0.139      | 0.018     |            | (<0.01)      |
|           |         | $\theta_2$ | 0.756      | $<0.001$  |            |              |
| Surabaya  | (0,1,2) | $\theta_2$ | 0.392      | $<0.001$  | Yes        | 0.08         |
|           |         | $\theta_3$ | 0.133      | 0.048     |            | (<0.01)      |
|           |         | $\theta_1$ | 0.724      | $<0.001$  |            |              |
| Surakarta | (0,1,2) | $\theta_2$ | 0.344      | $<0.001$  | Yes        | 0.10         |
|           |         | $\theta_3$ | 0.265      | $<0.001$  |            | (<0.01)      |
|           |         | $\theta_1$ | 0.758      | $<0.001$  |            |              |

![Figure 1](image1)

![Figure 2](image2)
Moreover, the results at Table 1 show that ARIMA model in each location could not fulfill the normal distribution assumption of the residuals caused by outliers. Therefore, it is necessary to handle these outliers by applying outlier detection. In this study, outlier detection on ARIMA modeling is implemented to obtain the ten most influential outliers.

3.2. VARIX and GSTARIX Model
As aforementioned, VARIX and GSTARIX modeling for the number of tourist arrivals data at four locations in Indonesia is done in two steps. The following are the results at each step of modeling process.

3.2.1. The first step: Results of TSR Model
The results of TSR model show that trend and seasonal pattern have significant impact on the forecast at the first step in 3 locations, i.e. Jakarta, Denpasar, and Surabaya. Otherwise, the results of TSR model in Surakarta show that only seasonal pattern has significant influence on the forecast. Moreover, the addition of ten outlier’s detection imply the forecast become more precise as actual data as shown at Figure 2. These ten outliers indicate some events have affected the number of tourist arrivals to some locations in Indonesia, such as monetary crisis on July 1997, bomb at Atrium Plaza on September 2001, and Bali bombing both on October 2002 and October 2005.

In addition, the TSR model shows that the event influencing negatively the number of tourist arrivals to Jakarta and Surabaya is only monetary crisis on July 1997. Whereas, some events affecting negatively the number of tourist arrivals to Denpasar are monetary crisis on July 1997, bomb at Atrium Plaza on September 2001, and Bali bombing both on October 2002 and October 2005. Moreover, number of tourist arrivals to Surakarta is affected by monetary crisis on July 1997 and Bali bombing II on October 2005. The graphics of forecast values at the first step both at in-sample and out-sample dataset are shown at Figure 3.

Figure 1. ACF and PACF Plot of Tourist Arrival Data after Non-Seasonal Differencing or $d=1$. 
Figure 2. Plot Time Series Actual Data with Time Series Regression (TSR) Model.

Figure 3. Time Series Plot between Actual and Forecast of TSR model

3.2.2. Second step: Result of VARI and GSTARI model

In this second step, residual of the first step is modeled by VARI and GSTAR models. The MCCF at Figure 4 shows that these residuals are not stationer in mean at non-seasonal lags. It is shown by many signs (+) and (-) on the MCCF schematic representation. Hence, differencing on non-seasonal order (d=1) is needed to make stationary data and the result of the MCCF is shown at Figure 5.

The MCCF of residual after implementing non-seasonal differencing shows that the data already satisfy stationary condition that illustrated by many signs (+) in this MCCF schematic representation. Then, MPCCF schematic representation and AIC values at some order VAR and VMA are needed to identify the best order of VAR model. The results are shown at Figure 6 and Table 2 for MPCCF and AIC, respectively.
The model, the GSTARI model has less.

As aforementioned, the forecast of three locations. As example, the best GSTARI model contains of insignifient parameters in temporal order. It implies the temporal order of GSTARI model is 3 with non-seasonal differencing, and the spatial order is determined following the first spatial order. Hence, the GSTARI model that be applied for these tourist data is GSTAR(3)-I(1). Several spatial weights are employed in this GSTARI model, such as OLS and GLS model, respectively.

Moreover, the GSTARI model is built by using the results of previous VAR model particularly about its temporal order. It implies the temporal order of GSTARI model is 3 with non-seasonal differencing, and the spatial order is determined following the first spatial order. Hence, the GSTARI model that be applied for these tourist data is GSTAR(3)-I(1). Several spatial weights are employed in this GSTARI model, such as uniform, inverse of distance, NCC, and NIPCC weights. Two estimation methods are used for estimating the parameters, i.e. OLS and GLS, so the models are known as GSTARI-OLS and GSTARI-GLS model, respectively.

Unlike the VAR model, the GSTARI model has less parameters than VAR model, i.e. 24 parameters. The insignificant parameters in GSTARI model by both OLS and GLS methods were also excluded from the model. Finally, the best GSTARI model contains of 11 temporal parameters and 1 spatial parameter. Moreover, the results of GSTARI model show that only number of tourist arrivals to Surakarta has relationship with other three locations. As example, the final GSTAR-OLS(3)-I(1) model with uniform weight can be written as:

\[
\begin{bmatrix}
N_t^1(t) \\
N_t^2(t) \\
N_t^3(t) \\
N_t^4(t)
\end{bmatrix}
= 
\begin{bmatrix}
-0.589 & 0 & 0 & 0 \\
-0.583 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
N_{t-1}^1 \\
N_{t-1}^2 \\
N_{t-1}^3 \\
N_{t-1}^4
\end{bmatrix}
+ 
\begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
N_{t-3}^1 \\
N_{t-3}^2 \\
N_{t-3}^3 \\
N_{t-3}^4
\end{bmatrix}
+ 
\begin{bmatrix}
0.589 & 0 & 0 & 0 \\
0.583 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\epsilon_t(t) \\
\epsilon_t(t) \\
\epsilon_t(t) \\
\epsilon_t(t)
\end{bmatrix}
\]

Table 2. AIC Value of Residual After Differencing Lag 1

| Lag | MA(0) | MA(1) | MA(2) | MA(3) | MA(4) | MA(5) |
|-----|-------|-------|-------|-------|-------|-------|
| AR(0) | 63.4213 | 62.5120 | 62.5383 | 62.6039 | 62.6264 | 62.6669 |
| AR(1) | 62.7429 | 62.5999 | 62.5978 | 62.6509 | 62.7162 | 62.7427 |
| AR(2) | 62.6514 | 62.6140 | 62.6583 | 62.6582 | 62.7497 | 62.7877 |
| AR(3) | **62.6114** | 62.6307 | 62.7191 | 62.7461 | 62.8111 | 62.8985 |
| AR(4) | 62.6743 | 62.7163 | 62.7714 | 62.8413 | 62.8779 | 62.9478 |
| AR(5) | 62.7330 | 62.7732 | 62.8606 | 62.9670 | 63.0325 | 64.2125 |

The results of MPCCF show that the significant lags are 1, 2, and 3, whereas the smallest AIC is for AR(3). It means that the appropriate order of model based on both statistics is VAR(3)-I(1). This model yields 48 parameters, i.e. 16 parameters for each lag 1, 2, and 3. The estimation and significance test show that not all parameters are significant. Then, these insignificant parameters are eliminated from the model by applying restriction in backward scheme. The final model consists of 11 parameters and satisfies white noise assumption of the residuals. Thus, the best subset VAR([3])-I(1) model can be written as:

3.2.3. The forecast accuracy comparison

As aforementioned, the forecast of VARIX or GSTARIX model is a combination of forecast from TSR model in the first level and VARIOR GSTARI model in the second level, respectively. The comparison results of RMSE at out-sample dataset are shown in Table 3.
Table 3. The results of RMSE comparison at out-sample dataset

| Model and Weight | Jakarta | Denpasar | Surabaya | Surakarta |
|------------------|---------|----------|----------|-----------|
| VARIX            | 40,500  | 44,890   | 2,778    | 412       |
| GSTARIX-OLS      |         |          |          |           |
| Uniform          | 40,414  | 44,915   | 2,761    | 413       |
| Inverse of distance | 40,414   | 44,915   | 2,761    | 415       |
| NCC              | 40,414  | 44,915   | 2,761    | 400       |
| NIPCC            | 40,414  | 44,915   | 2,761    | 413       |
| GSTARIX-GLS      |         |          |          |           |
| Uniform          | 40,436  | 44,906   | 2,776    | 411       |
| Inverse of distance | 40,432   | 44,908   | 2,776    | 414       |
| NCC              | 40,480  | 44,899   | 2,776    | 398       |
| NIPCC            | 40,436  | 44,906   | 2,776    | 411       |

Note: Bold number shows the smallest values of RMSE

The results on Table 3 show that the best model for the number of tourist arrivals to Jakarta and Surabaya is GSTARIX-OLS with all types of weights, such as uniform, inverse of distance, NCC, and NIPCC weights. Otherwise, the best models for data in Denpasar and Surakarta are VARIX and GSTARIX-GLS model with NCC weight, respectively.

Figure 7. Comparison of Forecasting Effectiveness Based on RMSE Out-sample Data.

Moreover, the VARIX and GSTARIX models produce forecast values with almost similar accuracy. Figure 7 shows the RMSE k-step forecast of the VARIX and GSTARIX models for predicting the number of tourist arrivals at four areas in Java and Bali. These graphs illustrate that VARIX and GSTARIX models yield consistent and accurate forecast until 6-step-ahead prediction. It is supported by the RMSE k-step-ahead that tend to have low and consistent value till 6 steps ahead in almost four locations of tourist arrivals in Java and Bali.

4. Conclusion

Based on the result and discussion at the aforementioned section, it could be concluded that the proposed two-level of VARIX and GSTARIX models could work well to reconstruct the trend, seasonal, and outlier detection at spatio-temporal data. The first level employed TSR model for handling trend, seasonal, and outliers effect, and the second level applied VARI or GSTARI model for tackling spatio-temporal dependency.

Due to the data pattern for each location is difference particularly about the impact of outliers, the best spatio-temporal models for forecasting the number of tourist arrivals to four locations in Java and Bali are not unique. In general, GSTARIX-OLS is the best model for forecasting tourist arrivals to Jakarta and Surabaya, GSTARIX-GLS for Surakarta case, and VARIX for forecasting tourist arrivals...
to Denpasar. Moreover, the spatial relationship existed only in tourist arrivals to Surakarta. Additionally, Further research is needed to validate the results of this comparison study particularly by applying to other more spatio-temporal data.

5. Acknowledgment

This study was funded by DRPM under “Penelitian Dasar Unggulan Perguruan Tinggi (PDUPT)” scheme No. 930/PKS/ITS/2018. The authors thank to the General Director of DIKTI for supporting and to the reviewers for the valuable suggestions.

6. Reference

[1] P Pfeifer and S Deutsch 1980 “A Three Stage Iterative Procedure For Space-Time Modeling Technometrics” 22(1) pp 35-47.
[2] S A Borovkova, H P Lopuhaa, and B N Ruchjana 2002 “Generalized STAR with Random Weights, Proceeding of the 17th International Workshop on Statistical Modeling”, Chania-Greece, pp 143-151.
[3] B N Ruchjana, A Generalized 2002 “Space-Time Autoregressive Model and Its Application to Oil Production”, Unpublished Ph.D Thesis, Institut Teknologi Bandung.
[4] Suhartono and R. M. Atok 2005 “Comparison between VARIMA and GSTAR Model for Forecating Space-Time Data, National Seminar of Statistics”, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.
[5] Suhartono and Subanar 2006 “The Optimal Determination of Space Weightin GSTAR Model by UsingCross-CorrelationInference”, Journal of Quantitative Methods, 2(2), pp 45-53.
[6] D U Wutsqa, and Suhartono 2010 “Seasonal Multivariate Time Series Forecasting on Tourism Data by using VAR-GSTAR Model”, Jurnal Ilmu Dasar, 11(1), pp 101-109.
[7] Setiawan, Suhartono and M. Prastuti 2016 “S-GSTAR-SUR Model for Seasonal Spatio-Temporal Data Forecasting”, Malaysian Journal of Mathematical Sciences, 10, pp 53-65.
[8] S C Nisak 2016 “Seemingly Unrelated Regression Approach for GSTARIMA Model to Forecast Rain Fall Data in Malang Southern Region Districts”, CAUCHY, 4(2), pp 57-64.
[9] J H P Messakh, M N Aidi, and F M Afendi 2017 “Rice Harvest Area Modeling with GSTARIMA on Six Provinces in Indonesia”, International Journal of Scientific & Engineering Research, 8(8), pp 424-427.
[10] H Bonar, B N Ruchjana, and G Darmawan 2017 “Development of Generalized Space Time Autoregressive Integrated with ARCH error (GSTARI–ARCH) model based on consumer price index phenomenon at several cities in North Sumatera province”. In AIP Conference Proceedings (Vol. 1827, No. 1, p. 020009), AIP Publishing.
[11] A P Ditago, and Suhartono 2015 “Simulation Study of Parameter Estimation Two-Level GSTARX-GLS Model”, International Seminar on Sciences and Technology, pp 167-168.
[12] R Mubarak, and Suhartono 2016 “Simulation of Generalized Space-Time Autoregressive with Exogenous Variables Model with X Variable of Type Metric”, IPTEK Journal of Proceedings Series, 2(1), pp 173-174.
[13] Suhartono, S R Wahyuningrum, Setiawan, and M. S. Akbar 2016 “GSTARX-GLS Model for Spatio-Temporal Data Forecasting”, Malaysian Journal of Mathematical Sciences, 10, pp 91-103.
[14] W W S Wei 2006 “Time Series Analysis Univariate and Multivariate Methods, 2nd edition”, USA: Pearson Education Inc.
[15] W Greene 2002 Econometric Analysis, 5th edition, New Jersey: Prentice Hall.
[16] V Srivastava, and T Dwivedi 1979 “Estimation of Seemingly Unrelated Regression Equation, A Brief Survey”, Journal of Econometrics, pp 15-32.