

# Prediction and Simulation Spatio-Temporal Support Vector Regression for Nonlinear Data

Khalilur Rahman <sup>1</sup>, Margaretha Ari Anggorowati<sup>1</sup>, Agung Andiojaya<sup>1</sup>

<sup>1</sup>Badan Pusat Statistik

*Jl. Dr Sutomo No 6-8, Jakarta Pusat*

\*[khalilur.rahman@bps.go.id](mailto:khalilur.rahman@bps.go.id)

## Abstract

Spatio-temporal model forecasting method is a forecasting model that combines forecasting with a function of time and space. This method is expected to be able to answer the challenge to produce more accurate and representative forecasting. Using the ability of method Support Vector Regression in dealing with data that is mostly patterned non-linear premises n adding a spatial element in the model of forecasting in the form of a model forecasting Spatio- Temporal. Some simulations have done with generating data that follows the Threshold Autoregressive model. The models are correlated into spatial points generated by several sampling methods. Simulation models are generated to comparing the accuracy between model Spatio-Temporal Support Vector Regression and model ARIMA based on Mean Error, Mean Average Error, Root Mean Square Error, and Mean Average Percentage Error. Based on the evaluation results, it is shown that forecasting with the Spatio-Temporal Support Vector Regression model has better accuracy than forecasting ARIMA.

**Keywords:** Nonlinear, Threshold Autoregressive, Spatio-Temporal, Support Vector Regression

## I. INTRODUCTION

Forecasting becomes an important thing in the decision making and planning process. According to Supranto (2008), forecasting is a conjecture or estimate regarding the occurrence of an event or event in the future. One of the objectives of forecasting, especially in the economic field, is an early warning of economic turmoil that will occur, but it can also be used as a picture of the economic situation in the future. With accurate forecasting, it is expected to be a solution for the creation of national economic stability.

The growing need for forecasting, demands a forecasting method is also able to handle problems based on area. This is because forecasting that generally uses data that has a time function as in the time series model is felt to be still less representative and accurate because it does not accommodate the regional factors that occur. Therefore, it is necessary to integrate time series data with spatial data to form a spatio-temporal data used for forecasting.

According to Kamarianakis (2001), spatio-temporal data is a time-series data that appears in spatial processes. The use of spatio-temporal data, which includes spatial data in time series data, can be a solution in area and time-based forecasting. This is because forecasting with spatial temporal data considers the spatial influence between dependent observations and also non-dependent observations that form time-series patterns. In other words, this forecasting considers spatial correlation and time-series correlation. This is expected to be a solution in forecasting time and space-based forecasting.

Time and space-based forecasting with spatio-temporal data has been done in various studies. Some of these studies include: forecasting fires on peat lands in Riau Province (Kirono, 2016), forecasting the Consumer Price index in the foodstuffs group (Hapsari, 2017), forecasting rainfall in Jember Regency (Retnaningrum, 2015), forecasting past flows cross-residents (Xu, Wang, & Wang, 2014), forecasting temperatures in the city of Guangzhou (Cheng, Wang, & Li, 2007), and forecasting the performance of irrigation operating systems on the North Plate river (Mohan, 2014).

Some data relating to economics has a non-linear pattern, which is difficult to forecast using conventional

forecasting methods. For this reason, forecasting methods are needed that are able to handle non-linear patterns, one of which is Support Vector Regression. Support Vector Regression, as a method non-parametric-based Machine Learning, is known to be able to handle the data pattern of non-linear. According to Vapnik (1995), Support Vector Regression is a Machine Learning method based on statistical learning theory that meets the principle of structural risk minimization with the aim of minimizing empirical risk and model complexity and providing high generalization capabilities. The use of Support Vector Regression to forecast non-linear Spatio-Temporal is done by Cheng (2010) to forecast the temperature of Guangzhou city, and Xu Ya Nyan (2014) to forecast population traffic flow.

Spatio-Temporal Support Vector Regression is strongly influenced by the kernel of the spatial and temporal kernel. Spatial kernels play a role in spatial factors and temporal kernels to handle temporal factors. The choice of spatial kernels and temporal kernels is an important thing in forecasting Spatio-Temporal Support Vector Regression. In this study, a combination of spatial kernels and temporal kernels will be evaluated in forecasting Spatio-Temporal Support Vector Regression. This study limits the model Temporal Linear Kernel, Radial Basis Function, and Polynomials, as well as Spatial Kernel in the form of Rectangular and Gaussian.

The purpose of this study is to simulate and predict the spatio-temporal Support Vector Regression method that can forecast the non-linear Spatio-temporal data. Methods of Spatio-Temporal Support Vector Regression will be evaluated and compared to its accuracy with the most popular time series statistical forecasting methods, ARIMA.

## II. LITERATURE REVIEW

### A. Time Series Modelling

The Time Series model forecasting method is a method of forecasting using patterns of past data events to predict future event patterns. The formation of a time series model for forecasting is obtained from the information contained in the data series. One time series model that is quite often used is the ARIMA Box-Jenkins model. The ARIMA model is a time series model that uses information in its series to produce models. The ARIMA model is basically a combined model of AR (Autoregressive) and MA (Moving Average). The AR model uses the autocorrelation information of current observations to past observations (Autoregressive), while the MA model uses information on past deviations (Moving Average). The Box-Jenkins ARIMA method is stated in equation (1) below:

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} \quad (1)$$

$Y_t$  = Time observation value (t)

$Y_{t-i}$  = Time-observed value (ti)

$e_{t-1}$  = time observation deviation (tj)

C = intercepts

$\phi$  = Autoregressive parameter

$\theta$  = Moving Average parameter

### B. Autocorrelation

An autocorrelation is a form of relationship (correlation) that occurs in a time series (time series) with a lagged version of the time series itself, which is also called temporal autocorrelation. Temporal autocorrelation is used to measure and explain the internal relationship between observations in a time series. In other words, temporal autocorrelation is used to explain the relationship of the value of the current variable with the value of the variable in the past.

Calculation of autocorrelation, clicking generate value ranging from +1 to -1. +1 autocorrelation represents a perfect positive correlation (the increase seen in one time series leads to a proportional increase in another time series). In contrast, autocorrelation -1 represents a perfect negative correlation (an increase seen in one time series results in a proportional decrease in another time series).

### C. Spatial Modelling

The spatial model is modeling the relationship between the dependent variable and the independent variable, which is influenced by the weight of the location of the independent variable. Spatial modeling is stated in equation (2) below:

$$y_i = \beta_0 (u_i, v_i) + \sum_{k=1}^n x_{ik} \beta_k(x_i, v_i) + \varepsilon \quad (2)$$

$y_i$  = dependent variable

$x_i$  = independent variable

$(u_i, v_i)$  = coordinate value

$\beta_k(x_i, v_i)$  = spatial weighting function

In spatial modeling known as spatial weighting, the spatial weighting is carried out because of the different spatial effects at each location. This weighting is done to improve the accuracy of the spatial forecasting model by accommodating the spatial autocorrelation that occurs.

#### D. Spatial Autocorrelation

The concept of spatial autocorrelation is an extension of temporal autocorrelation. In temporal autocorrelation, the observed object is the time that runs in one direction and one dimension. In the spatial autocorrelation, the object observed is an area (spatial), which has two dimensions and is complex in shape.

Spatial autocorrelation values are used to describe the relationship of values between spatial locations with each other. Spatial autocorrelation values are used to observe the pattern of values of a spatial location influenced by the value of the surrounding spatial location.

#### E. Spatio-Temporal Modelling

Spatio-Temporal Model is basically a combination of a time series model which has temporal autocorrelation and also spatial models that have spatial autocorrelation. Spatio-Temporal forecasting method is generally a forecast that combines temporal forecasting functions and spatial forecasting functions into a non-linear forecasting combination function. The non-linear combination function is expressed as equation (3):

$$\hat{y} = \phi(X) = \phi(f_t, f_s) \quad (3)$$

$X$  = attribute vector that consist of  $f_t$  (temporal function) and  $f_s$  (spatial function)

$\phi(X)$  = nonlinier prediction function

$\hat{y}$  = predicted value

#### F. Support Vector Regression

One effective machine learning method in forecasting is Support Vector Regression (SVR). SVR is a development of the machine learning method Support Vector Machine (SVM), which was introduced by Vapnik (1995) and is effective in dealing with regression (SVR), especially in non-linear data. According to Kanevski (2000) and Cheng et.al(2007), SVR has several advantages that become interesting characteristics, including:

- Has a robust solution (resistant) to noise,
- Sparseness regression is the ability to produce accurate results even with the number of observations that are rare or small,
- Automatic control of the complexity of solutions,
- Good generalization performance (estimation accuracy).

#### G. Spatio-Temporal Support Vector Regression

The Spatio-Temporal Model stated in equation (3) is a linear combination of spatial functions and temporal functions. Spatial functions and temporal functions in the Spatio-Temporal Support Vector Regression can be interpreted as a linear combination of the Temporal Kernels function ( $K_T$ ) and Spatial Kernels ( $K_S$ ). The results of these linear combinations which then form the Spatio-Temporal Kernels ( $K_{ST}$ ).

#### H. Temporal Kernels

Temporal Kernels function is used to do time series forecasting, which is able to handle data nonlinearity problems that occur in time functions. Some kernels functions used in this study are Linear Kernels and Radial Kernels Basis. The three functions are formulated as equality (4), and (5) as follows:

Kernels Linear Function

$$K_T(X, X') = \langle X, X' \rangle \quad (4)$$

Radial Basis Kernels Function is given by

$$K \langle X, X' \rangle = \exp \left( -\frac{1}{2\sigma^2} |X - X'|^2 \right) \quad (5)$$

where  $\sigma$  denotes the deviation of training data taken as a support vector.

#### I. Spatial Kernels

Function *Spatial Kernels* is a function of the distance that is used to generate a spatial weighting matrix function on the model *Spatio-Temporal*. The *Spatial Kernels* function used in this study is the *Gaussian Kernels* function. *Spatial Kernels* functions are formulated in equations (6).

$$K_S \left( \frac{d_{ij}}{h} \right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{d_{ij}}{h} \right)^2} \quad (6)$$

#### J. Spatio-Temporal Kernels

Function *Spatio-Temporal Kernels* on a *Support Vector Regression* is a linear combination of the functions *kernels* spatial and function *kernels* temporal. The linear combination function is formulated in equation (7) as follows:

$$K_{ST}(\dots) = \sum_{i=1}^N (\alpha_i^* - \alpha_i) (K_T(\dots), K_S(\dots)) \quad (7)$$

$\alpha_i$  atau  $\alpha_i^*$  = support vector not zero at non-linear prediction function

$K_{ST}(\dots)$  = Spatio-Temporal kernel functions

#### K. Non-linear Model

Non-linear model is a model that shows the relationship between variables not in a linear or straight line form. Some non-linear models in the time series model are the Threshold Autoregressive (TAR) model. Model Threshold Autoregressive is a model time series non-linear that was segmented so that among the different segments have different models. The TAR model was first proposed by Tong (1983, 1990).

TAR (Threshold Autoregressive) model is formulated in equation (10):

$$y_t = \begin{cases} \alpha_{y_{t-1}} + \varepsilon_{L_t} & \text{if } y_{t-1} \leq r \\ \beta_{y_{t-1}} + \varepsilon_{U_t} & \text{if } y_{t-1} > r \end{cases} \quad (10)$$

$y_t$  = time series data 1, 2, ... T,

$\alpha$  = AR coefficient in the lower regime,

$\beta$  = AR coefficient in the upper regime

$\varepsilon_{L_t}$  = residual in lower regime,

$\varepsilon_{U_t}$  = residual in upper regime

### III. RESEARCH METHOD

#### A. Data Simulation Method

In this study, a simulation study was carried out using data generation. In the simulation study, it was generated using two non-linear time series models, namely: the Threshold Autoregressive Model. Data generated as many as 16 columns with each column having 720 series. Data time series formed later related by 16 Spatial points were created with multiple sampling methods in this study. Data generated with generating function was determined several plant parameters as follows:

TABLE I  
GENERATING FUNCTION PARAMETER IN AUTOREGRESSIVE THRESHOLD MODEL

Parameter	Description	Value
(1)	(2)	(3)
$\alpha$	Regime lower coefficient	1
$\beta$	Regime upper coefficient	1
R	Regime	5000
e	Residual regime lower	100
Initial value	Start value	10000
Burn-in value	Burn-in Value	100
N	Series long	720

TABLE II  
GENERATING FUNCTION PARAMETER IN SPATIAL POINTS

Parameter	Description	Value
(1)	(2)	(3)
$\alpha$	Regime lower coefficient	1
$\beta$	Regime upper coefficient	1
R	Regime	5000
e	Residual regime lower	100
Initial value	Start value	10000
Burn-in value	Burn-in Value	100
N	Series long	720

B. Analysis Method

The analytical method used consists of exploratory data analysis and *Spatio-Temporal* forecasting analysis. Data exploration analysis is used to find temporal and spatial patterns in the data while forecasting analysis is used to forecast the *Spatio-Temporal* data. In forecasting analysis *Spatio-Temporal Support Vector Regression* do forecasting using six combinations of the temporal and spatial kernel as follows:

TABLE III  
SPATIO-TEMPORAL SUPPORT VECTOR REGRESSION PARAMETER MODEL

No	Kernel Temporal	Kernel Spatial
(1)	(2)	(3)
1	Linear (C = 1, epsilon = 0,1)	Rectangular (d = great circle distance)
2	Radial Basis Function (C = 1, epsilon = 0,1)	Rectangular (d = great circle distance)

IV. RESULTS AND DISCUSSION

A. Exploration Data Analysis and Simulation

Simulation studies conducted in this research are used to validate the development of algorithms that are built. Simulation studies use generated data with non-linear functions. The generated data is built according to the Threshold Autoregressive model. There are 16 models that have been raised are then mapped into 16 shape points selected with a particular sampling method.

Generated data that has been built is then carried out an analysis of data exploration to see the characteristics of the data generated, both in terms of time series and its spatial side. Exploration data analysis is done by visualizing the time series data to the time series line graph. Following is the time series line graph of the generated data model:

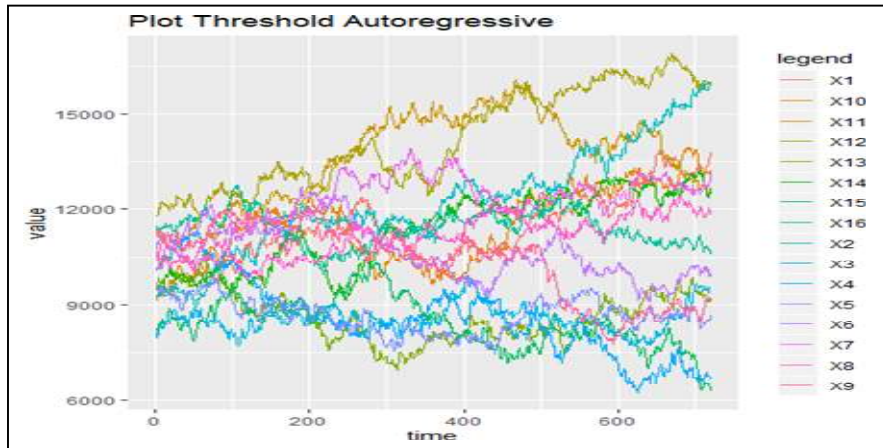


Fig. 1. Time Series Plot of Data Simulation Model Threshold Autoregressive

Spatial exploration analysis will be mapping spatial points that have been sampled using several sampling methods into polygon form. The following are spatial points obtained using several sampling methods that present areas of an area:

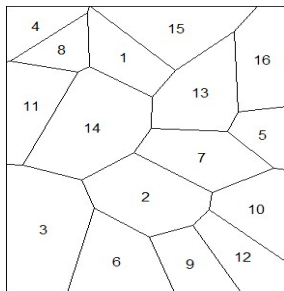


Fig. 2. Clustered Sampling Spatial Points

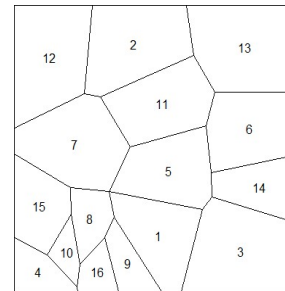


Fig. 3. Random Sampling Spatial Points

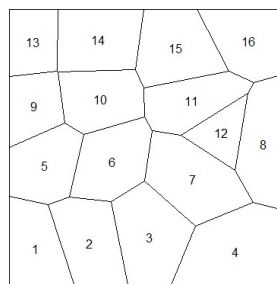


Fig. 4. Random Sampling Spatial Points

Spatio-temporal exploration analysis will be mapping each of the generated data values from the time series model into spatial points in the form of a spatio-temporal graph which is a choropleth map containing the change of color in the polygon area at each time interval. The color in the dark blue polygon area indicates the

lower generation data value, while the yellow polygon area indicates the higher rise data value. Here is the spatio-temporal graph of the generated data model:

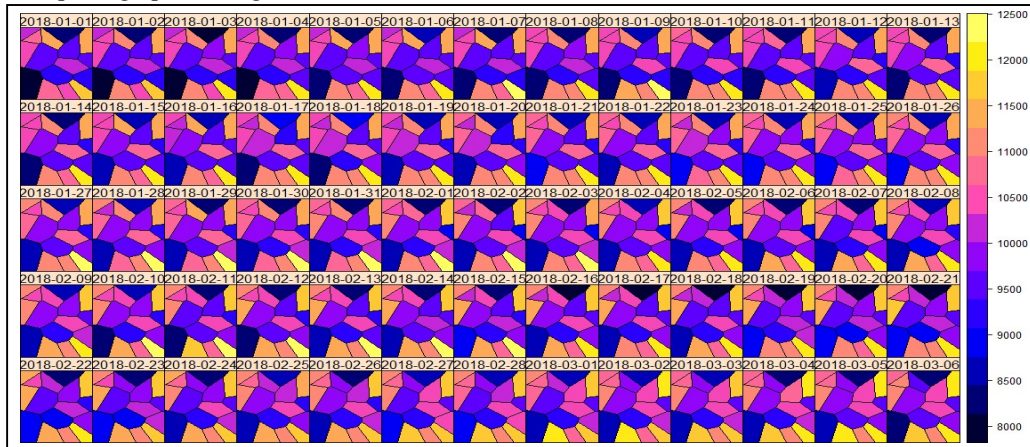


Fig. 5. Spatio-temporal Plot of Data Simulation Model Threshold Autoregressive in Clustered Spatial Points

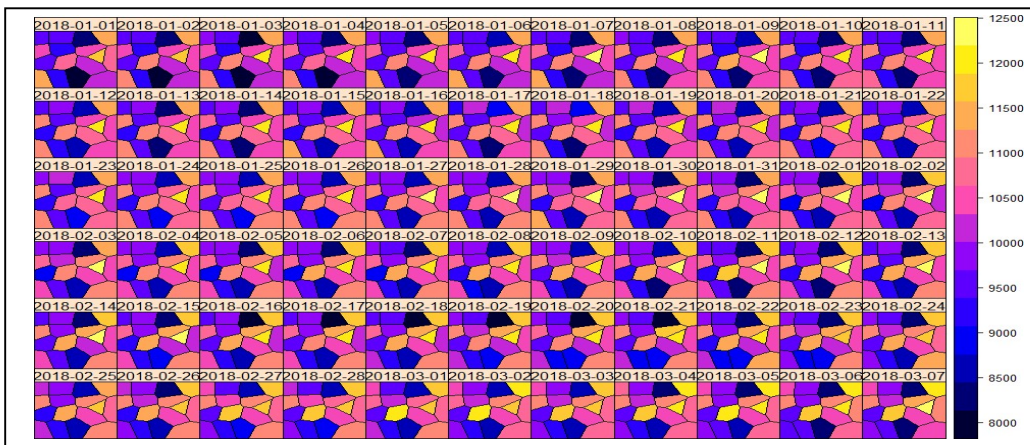


Fig. 6. Spatio-temporal Plot of Data Simulation Model Threshold Autoregressive in Stratified Spatial Points

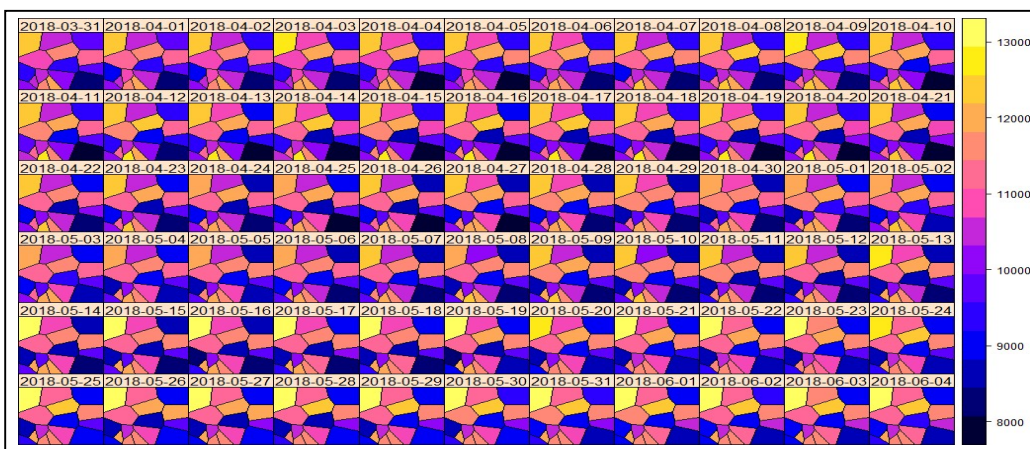
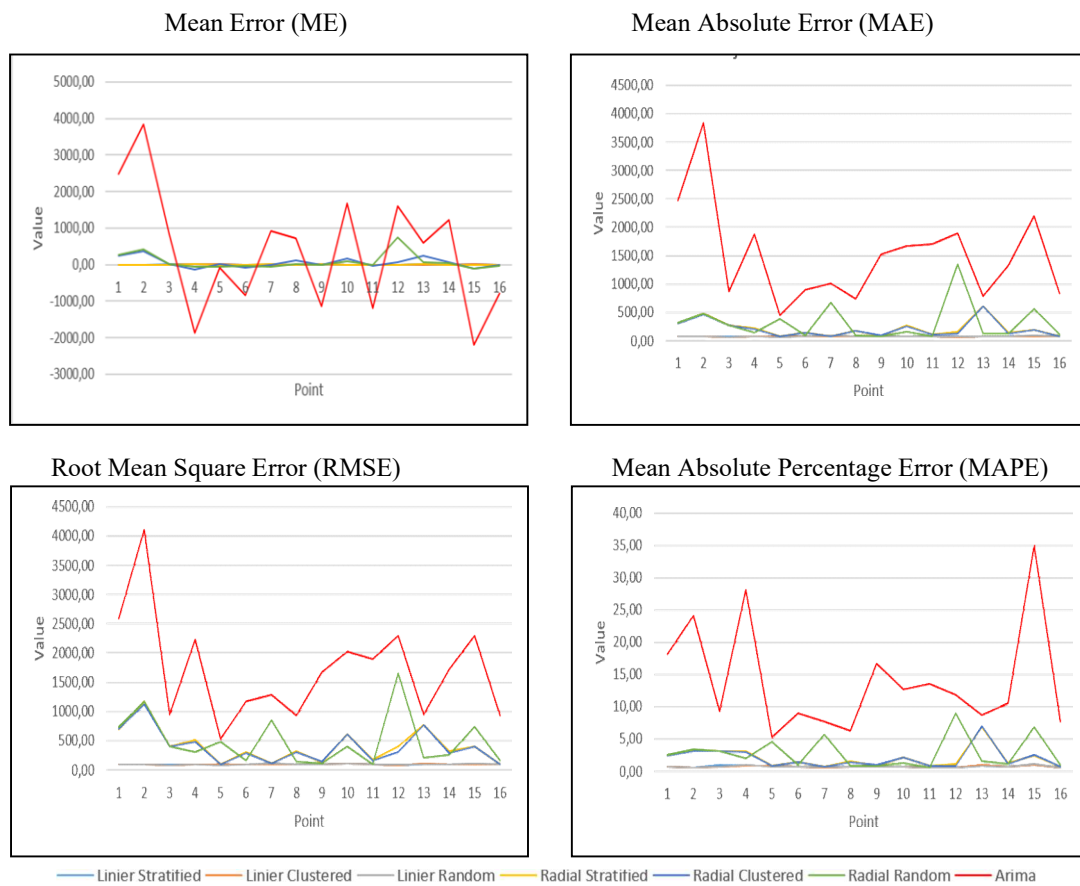


Fig. 6. Spatio-temporal Plot of Data Simulation Model Threshold Autoregressive in Random Spatial Points

*B. Spatio-Temporal Analysis*

Based on the results of generating data that have been carried out using Data Exploration Analysis. So Spatio- Temporal Support Vector Regression forecasting is performed for each combination of the generated data model and spatial points using various combinations of spatial kernels and temporal kernels that have been determined.

The use of spatial kernels in Spatio-Temporal Support Vector Regression is used to form spatial weight while temporal kernels are used to forecast time series. The following is a comparison of the accuracy of the results of forecasting Spatio-Temporal Support Vector Regression:



Picture 8. Accuracy Test Result

Based on the results of the accuracy testing of the Spatio-Temporal Support Vector Regression model for Threshold Autoregressive generation data, it can be seen that there is a significant difference in accuracy in forecasting results using the Spatio-Temporal Support Vector Regression when compared to forecasting made with traditional forecasting such as ARIMA. Among the Spatio-Temporal Support Vector Regression forecasting models, the linear-gaussian kernel is able to produce more accurate forecasting results with ME, MAE, RMSE, and MAPE values that are small close to 0.



## V. Conclusion

The development of the Spatio-Temporal Model is intended to answer the demand for forecasting models that not only depend on the effects of time but are also able to accommodate changes because regional dependence, it brings to the development of Spatio-Temporal forecasting methods. The Spatio-Temporal Support Vector Regression forecasting model is a kind of Spatio-Temporal forecasting methods that answered the problem, also demands in real data that have a non-linear pattern. The development of the Spatio-Temporal Support Vector Regression model by using a combination of Spatial Kernel and Temporal Kernel shows better accuracy results than the ARIMA model. This can prove that the Spatio-Temporal Support Vector Regression model can be a more effective forecasting method in forecasting non-linear patterned data and able to accommodate spatial effects and time effects. It is hoped that in the future, there will be more developments in Spatio-Temporal Support Vector Regression models with a combination of kernel-spatial and other temporal kernels to examine other non-linear forecasting models.

## ACKNOWLEDGMENT

The authors would thank our institution BPS-Statistics, who provided facilities for this research and support us doing this research. We would also like to show our gratitude to references for sharing their research so we can completely finish this research, and we thank my colleague friend Graham, who help us while doing this research. Finally, with all of our respect, we hope this research can give the real impact in the real world.

## REFERENCES

- Abu Awad, Y., Koutrakis, P., Coull, B. A., & Schwartz, J. (2017). A spatio-temporal prediction model based on support vector machine regression: Ambient Black Carbon in three New England States. *Environmental Research*, 159 (Agustus), 427–434.
- Borst, Richard. (2013). A Space-Time Model for Computer Assisted Mass Appraisal. *Aestimium*. 10.13128/Aestimium-13160.
- Cheng, T., & Wang, J. (2006). Applications of spatio-temporal data mining and knowledge discovery (STDMKD) for forest fire prevention. *Proc. of Multitemporal Data and Change Detection*,
- Cheng, T., Wang, J., & Li, X. (2007). The Support Vector Machine for Nonlinear Spatio-Temporal Regression. *Proceedings of Geocomputation 2007*, 1–6.
- Kamarianakis, Y. (2003). Spatial-Time Series Modeling: a Review of the Proposed Methodologies. *Phone Fax*, (217217), 61801–63671.
- Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab – An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, 11(9), 1–20.
- Kirono, S. (2016). Spatio-Temporal Sequential Pattern Mining Untuk Deteksi Dini Kebakaran Pada Lahan Gambut Di Provinsi Riau.
- Makridakis, S. G., Wheelwright, S. C., & Hyndman, R. J. (1997). *Forecasting Methods and Applications* (3rd ed.). New York: Wiley.
- Mohan, A. (2014). A New Spatio-Temporal Data Mining Method and its Application to Reservoir System Operation.
- Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support vector machine: Teori dan Aplikasinya dalam Bioinformatika. *IlmuKomputer.Com*.
- Ojemakinde, B. T. (2006). Support Vector Regression for Non-Stationary Time Series.

- Retnaningrum. (2015). Penerapan Model Star (Space Time Autoregressive) dan Arima (Autoregressive Integrated Moving Average) untuk Peramalan Data Curah Hujan Di Kabupaten Jember [Skripsi]. Jember: Universitas Jember.
- Smith, T. E. (2016). *Notebook on Spatial Data Analysis* [online].
- Supranto, J. (2008). Statistik: Teori dan aplikasi (7th ed., Vol. 1). Jakarta: Erlangga.
- Tong, H. (1983). Threshold models in non-linear time series analysis. New York: Springer
- Vapnik, V. N. (2000). The nature of statistical learning theory (2nd ed.). New York, USA: Springer-verlag.
- Wang, J., Cheng, T., & Li, X. (2007). Non-linear Integration of Spatial and Temporal Forecasting by Support Vector Machines. Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), (Fskd), 61–66.
- Wang, J. Q., Cheng, T., & Haworth, J. (2012). Space-Time Kernels. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 38(2), 57–62.
- Wikle, C. K. (2015). Modern perspectives on statistics for spatio-temporal data. Wiley Interdisciplinary Reviews: Computational Statistics, 7(1), 86–98.
- Xu, Y., Wang, B., & Wang, F. (2014). Spatio-temporal Variable Selection Based Support Vector Regression for Urban Traffic Flow Prediction.