Precipitation forecasting using neural network model approach

To cite this article: A Iriany et al 2020 IOP Conf. Ser.: Earth Environ. Sci. 458 012020

View the article online for updates and enhancements.
Precipitation forecasting using neural network model approach

A Iriany¹, D Rosyida¹, A D Sulistyono² and B N Ruchjana³

¹Department of Statistics, Faculty of Mathematics and Natural Science, Brawijaya University, Indonesia.
²Faculty of Fisheries and Marine Science, Brawijaya University, Indonesia.
³Department of Mathematics, Faculty of Mathematics and Natural Science, University of Padjajaran, Bandung, Indonesia.

*E-mail: atiekiriany@yahoo.com

Abstract. Neural network constitutes a non-linear model requiring no statistical assumption. Along the development of which, neural network model has been frequently combined with time series and spatio temporal models. This current research combined neural network and spatio temporal models. One of spatio temporal models is GSTAR-SUR model. The weight projected in this current research is cross covariance normalized weight. This sort of weight is deemed suitable for data with high variability. The significant variable in GSTAR-SUR model containing cross covariance normalized weight was used as input layer of neural network model. The hidden layer made use of 10 neurons fulfilling the criteria of the lowest RMSE value and there was 1 neuron used as output. The data were in the form of 10-day precipitations in Junggo, Pujon, Tinjumoyo, and Ngujung, during the period of 2005 to 2014. This research has found out that NN-GSTAR-SUR model yielded better and more accurate forecasting, showing $R^2$ value of 61.77%.

1. Introduction

Neural network is perceived as a universal model as it does not require any statistical assumption [1]. This method is adapted from the neural system of human brain [2]. Neural network is a non-linear model that has been much developed until recently. Some researchers have proven that neural network is salient and potent to solve a number of problems in various disciplines; one of whom is Zhang et al. [1] and Suhartono [3]. Neural network model has been frequently applied for non-linear time series. Some researchers applying neural network model for predicting time series are among others Suhartono and Endharta [4], Vlahogianni and Karlaftis [5] and Ostermark [6].

One of the components of neural network model is input layer. Input layer used in neural network model is said to be more flexible, including the one on spatio temporal data. Input layer could be in the forms of either raw or processed data resulted from former model. Suhartono [3] is the first scholar using input layer as that of time series in statistics. Since then, a lot more combinations between neural network and spatio temporal models have come to be under investigation; some of whom are Diani et al. [7] and Sulistyono et al. [8].

This current research combined neural network and spatio temporal methods. The research of Diani et al. [7] merely made use of one neuron in the hidden layer. Sulistyono et al. [8] chose to exclude weight component from input layer, though location weight is pivotal for spatio temporal model.
The renowned and developing spatio temporal model is Generalized Space Time Autoregressive (GSTAR), formerly introduced by Ruchjana [9]. Iriany and Ruchjana [10] further develop GSTAR model into GSTAR-SUR model. Location weight is an important component for GSTAR model. Location weights that have been recurrently used are uniform location weight, distance inverse, and cross covariance normalization [11,12]. However, those weights are improper for data with relatively high variability. Cross covariance normalization, in this current research, was used to detect relatively high data variability. This sort of weight was used as a component of input layer in neural network model.

2. Materials and Methods
The data, collected and analyzed in this current research, were in the form of 10-day precipitations during the period of 2005 to 2014. Those data were collected from Meteorological, Climatological, and Geophysical Agency of Karangploso. The precipitation data were representing some locations where majority of farmers reside in Malang Regency; the locations were in Junggo, Pujon, Tinjumoyo, and Ngujung.

The initial concept of this research was the determination of input layer. The determination of which was focusing on the formation of GSTAR-SUR model. It aimed at finding out significant variables of the model. The significant variables of GSTAR-SUR were then used as the input of neural network model. Diani et al. [7] assert that using significant variables on GSTAR model as the inputs of neural network model yields more accurate forecasting than using all existent variables.

GSTAR model is formed by using cross covariance normalized weight. This sort of weight has been researched and applied by Apanasovich and Genton (2010) to predict pollution in California. The calculation of weight is as follow:

$$W_{ij} = \frac{\text{cov}(z_{ij}(s))}{\sum_{k \neq s} \text{cov}(z_{ik}(s))}$$

$$\text{cov}(z_{ij}(s)) = r_{ij}(s) \sqrt{\left(\sum_{t=1}^{n} [z_i(t) - \bar{z}_i]^2\right) \left(\sum_{t=1}^{n} [z_j(t-s) - \bar{z}_j]^2\right)}$$

The parameter is estimated by employing Seemingly Unrelated Regression (SUR) with the following formula.

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

where $\Omega$ is variance covariance matrix of GSTAR model residuals resulted from Ordinary Least Square (OLS) [10]. In addition to input layer, the other components of neural network model are hidden layer and output layer. One hidden layer was used and there were of maximum 10 neurons used in the hidden layer. The determination of the number of neurons used in the hidden layer was based on the lowest RMSE value. One neuron was used as output layer [3]. Output layer is represented in vector, of which component is location variable as in $Y' = (Z_1 \ Z_2 \ Z_3 \ Z_4)'$. The used algorithm is the resilient backpropagation algorithm formerly used by Apriliyah et al [13] and Fadil et al [14] for estimating the sales of electricity load.

3. Results and Discussion
3.1. Data Identification
Precipitation is a vital factor for farmers in determining their planting patterns. Some locations in Malang Regency are mostly used as agricultural fields; they are Junggo, Pujon, Tinjumoyo, and Ngujung. The precipitations in the four locations were observed; there were in total 360 precipitation observations. The following stage of this research was forecasting the precipitations in those four locations. Each region has different precipitation intensity as shown in Table 1.
Table 1. Descriptive statistics of precipitations in four observed locations

<table>
<thead>
<tr>
<th>Location</th>
<th>Average</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junggo</td>
<td>6.248</td>
<td>6.954</td>
<td>0</td>
<td>35.800</td>
</tr>
<tr>
<td>Pujon</td>
<td>6.362</td>
<td>7.799</td>
<td>0</td>
<td>37.818</td>
</tr>
<tr>
<td>Tinjumoyo</td>
<td>4.948</td>
<td>5.683</td>
<td>0</td>
<td>29.000</td>
</tr>
<tr>
<td>Ngujung</td>
<td>5.012</td>
<td>5.902</td>
<td>0</td>
<td>35.600</td>
</tr>
</tbody>
</table>

The forecasting of precipitation was employing neural network model by means of input as that in GSTAR-SUR model. GSTAR-SUR modeling requires the data to be stationary against variance and average. The precipitation data were non-stationary against variance, but they were stationary against average, and thus it required Box-Cox transformation, that is \((Z_t + 2)^{-0.5}\) [15]. The stationary condition is to be fulfilled in order to find out significant parameter.

In GSTAR-SUR model, there is the component of location weight. The location weight, in this current research, was cross covariance weight. It was due to the relatively high variability of the precipitation data. This has been shown in Table 1 that standard deviation value is higher than the average of each location. The location weight of cross covariance normalization based on the stationary data is presented as follow:

\[
W = \begin{bmatrix}
0 & 0.3380 & 0.3243 & 0.3376 \\
0.3397 & 0 & 0.3296 & 0.3306 \\
0.3311 & 0.3338 & 0 & 0.3351 \\
0.3380 & 0.3355 & 0.3265 & 0
\end{bmatrix}
\]

The determination of time order in GSTAR-SUR model is based on MPACF and MACF as shown in Figure 1. The spatial order used is 1. Figure 1 shows that MPACF scheme cuts off up to lag 3; whereas MACF scheme is shown to form sine wave pattern. It means that the autoregressive order is said to be 3 and the moving average order is 0 [16]. Therefore, the formed order of GSTARX-SUR is GSTARX-SUR \((3_1)\).

Figure 1. MPACF (top) and MACF (bottom) schemes
3.2. **NN-GSTAR-SUR Model**

The formed neural network model was the one with the inputs of significant variables of GSTAR-SUR (3\(_1\)). This has been supported by a research conducted by Diani et al [7] finding that the inputs of significant variables in GSTAR-SUR model yield better forecasting than using all variables. Parameter forecasting was done by employing SUR model as GSTAR model constitutes multivariate model. SUR method potentially solves inter residual correlation [17]; accordingly, the parameter significance is more accurate compared to that in Ordinary Least Square (OLS) method.

Parameter significance test on GSTAR-SUR model has resulted in 16 significant variables out of 24 variables. These 16 variables were treated as input layer of neural network model. The number of neuron in the hidden layer was limited to range from 1 to 10 neurons and the best one was chosen based on RMSE value as shown in Table 2. The output layer had merely 1 neuron, in the form of vector.

<table>
<thead>
<tr>
<th>The Number of Neuron in Hidden Layer</th>
<th>RMSE Value</th>
<th>The Number of Neuron in Hidden Layer</th>
<th>RMSE Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1281</td>
<td>6</td>
<td>0.1227</td>
</tr>
<tr>
<td>2</td>
<td>0.1274</td>
<td>7</td>
<td>0.1207</td>
</tr>
<tr>
<td>3</td>
<td>0.1254</td>
<td>8</td>
<td>0.1172</td>
</tr>
<tr>
<td>4</td>
<td>0.1248</td>
<td>9</td>
<td>0.1178</td>
</tr>
<tr>
<td>5</td>
<td>0.1217</td>
<td>10</td>
<td>0.1162</td>
</tr>
</tbody>
</table>

The number of neurons used in the hidden layer was 10. It was due to the consideration that forecasting precipitation by means of 10 neurons in the hidden layer has yielded the lowest RMSE value, as shown in Table 2. Accordingly, the formed neural network model was NN (16,10,1) with the following architectural design.

![Diagram](image_url)

**Figure 2.** The best architectural design for NN (16,10,1) - GSTAR-SUR (3\(_1\)) model
Based on the architectural design in Figure 2, the equation of NN (16,10,1) - GSTAR-SUR (3) model is as below.

\[
\hat{Y}_{t,NN} = -0.578 + 2.715 f(h_1) + 1.239 f(h_2) + 0.229 f(h_3) + 1.317 f(h_4) \\
- 0.144 f(h_5) - 1.092 f(h_6) - 0.813 f(h_7) + 0.182 f(h_8) \\
- 0.429 f(h_9) - 1.034 f(h_{10})
\]

(3)

where \( f(h_i) \) is the activation function of logistic sigmoid in the hidden unit which is defined as follow:

\[
f(h_i) = \frac{1}{1 + e^{-h_i}}, \quad i = 1,2, \ldots, 10
\]

\( h_i \) was formed from variables as those in GSTAR (3) model. It includes the lag of precipitation data in each location and the multiplication between cross covariance location weight and the lag itself.

The following formula presents some components of \( h_i \) in lag 0.

\[
Z_{1t} = \begin{bmatrix} Y_{1t} \\ 0 \\ 0 \end{bmatrix}, \quad Z_{2t} = \begin{bmatrix} 0 \\ Y_{2t} \\ 0 \end{bmatrix}, \quad Z_{3t} = \begin{bmatrix} 0 \\ 0 \\ Y_{3t} \end{bmatrix}, \quad Z_{4t} = \begin{bmatrix} 0 \\ 0 \\ Y_{4t} \end{bmatrix}
\]

\[
F_{1t} = \begin{bmatrix} w_{12} Y_{2t} + w_{13} Y_{3t} + w_{14} Y_{4t} \\ 0 \\ 0 \end{bmatrix}, \quad F_{2t} = \begin{bmatrix} 0 \\ w_{21} Y_{1t} + w_{23} Y_{3t} + w_{24} Y_{4t} \\ 0 \end{bmatrix}
\]

\[
F_{3t} = \begin{bmatrix} 0 \\ 0 \\ w_{31} Y_{1t} + w_{32} Y_{2t} + w_{34} Y_{4t} \end{bmatrix}, \quad F_{4t} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\]

The followings are the neuron equations in the hidden layer.

\[
h_1 = -1.967 - 5.255 Z_{1,t-1} + 0.012 Z_{2,t-1} + 0.326 Z_{3,t-1} - 0.224 Z_{4,t-1} - 0.263 Z_{5,t-2} \\
- 1.706 Z_{6,t-2} + 1.857 Z_{7,t-3} + 0.525 Z_{8,t-3} - 1.203 F_{1,t-1} - 1.615 F_{2,t-1} \\
- 0.684 F_{3,t-1} + 1.573 F_{4,t-1} + 3.559 F_{5,t-2} - 1.494 F_{6,t-2} + 5.651 F_{7,t-3} \\
- 2.095 F_{8,t-3}
\]

\[
h_2 = 0.962 + 2.804 Z_{1,t-1} - 7.43 Z_{2,t-1} + 0.234 Z_{3,t-1} - 4.147 Z_{4,t-1} + 3.065 Z_{5,t-2} \\
+ 4.12 Z_{6,t-2} - 14.141 Z_{7,t-3} - 1.752 Z_{8,t-3} - 1.823 F_{1,t-1} + 34.943 F_{2,t-1} \\
+ 0.055 F_{3,t-1} + 1.4 F_{4,t-1} + 0.575 F_{5,t-2} + 13.534 F_{6,t-2} - 2.697 F_{7,t-3} \\
+ 1.427 F_{8,t-3}
\]

\[
h_{10} = 2.62 - 1.342 Z_{1,t-1} - 1.94 Z_{2,t-1} + 0.161 Z_{3,t-1} - 2.25 Z_{4,t-1} - 0.021 Z_{5,t-2} \\
+ 0.021 Z_{6,t-2} + 0.229 Z_{7,t-3} - 3.137 Z_{8,t-3} - 2.913 F_{1,t-1} - 5.986 F_{2,t-1} \\
- 2.859 F_{3,t-1} + 2.073 F_{4,t-1} + 0.263 F_{5,t-2} - 1.566 F_{6,t-2} - 2.412 F_{7,t-3} \\
- 0.515 F_{8,t-3}
\]

The reliability measurement of NN (16,19,1) – GSTAR-SUR (3) model as shown in equation (3) was seen from RMSE and R² values. The following Table 3 displays RMSE and R² values for each location.
Table 3. RMSE and $R^2$ values

<table>
<thead>
<tr>
<th>Location</th>
<th>RMSE Value</th>
<th>$R^2$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junggo</td>
<td>5.7830</td>
<td>0.5967</td>
</tr>
<tr>
<td>Pujon</td>
<td>5.8388</td>
<td>0.6412</td>
</tr>
<tr>
<td>Tinjumoyo</td>
<td>4.7784</td>
<td>0.5957</td>
</tr>
<tr>
<td>Ngujung</td>
<td>4.7486</td>
<td>0.6296</td>
</tr>
<tr>
<td>General</td>
<td>5.3131</td>
<td>0.6177</td>
</tr>
</tbody>
</table>

The results shown in Table 3 have imparted that NN (16,19,1) – GSTAR-SUR (31) model with cross covariance normalized weight is proper and suitable for forecasting precipitations in Junggo, Pujon, Tinjumoyo, and Ngujung. It has been proven by the $R^2$ value of 61.77%. In Pujon, the $R^2$ value has reached 64.12%. Figure 3 presents the comparison between the actual values of precipitation and the precipitation forecasting results in Junggo, Pujon, Tinjumoyo, and Ngujung within 6 months, from July to December 2014.

![Figure 3](image-url)

**Figure 3.** Precipitation forecast plots from July to December 2014 in several areas, such as Junggo (a), Pujon (b), Tinjumoyo (c), and Ngujung (d).

Referring to Figure 3, the forecasting results by means of NN (16,19,1) – GSTAR-SUR (31) model with cross covariance normalized weight have approached the actual data, though some dots are shown to be apart. However, the patterns resulted from this model are shown to resemble the actual data. It has
been proven that NN (16,19,1) – GSTAR-SUR (3) model with cross covariance normalized weight is reliable for forecasting precipitations in Junggo, Pujon, Tinjumoyo, and Ngujung.

4. Conclusion
The resulted model for forecasting precipitations in Junggo, Pujon, Tinjumoyo, and Ngujung is NN (16,19,1) – GSTAR-SUR (3). The cross covariance normalized weight on GSTAR-SUR model as input of neural network model has yielded better forecasting. This model has been found to show the highest $R^2$ value, reaching 61.77%. Accordingly, this model is said to be reliable.

References