ON HYBRIDIZATION OF TIME SERIES AND BAYESIAN REGULARIZATION NEURAL NETWORK MODELS

By

Jayson N. Payla and Bernadette F. Tubo, Ph.D

Presented by

Jayson N. Payla

Mindanao State University-Iligan Institute of Technology

Time Series

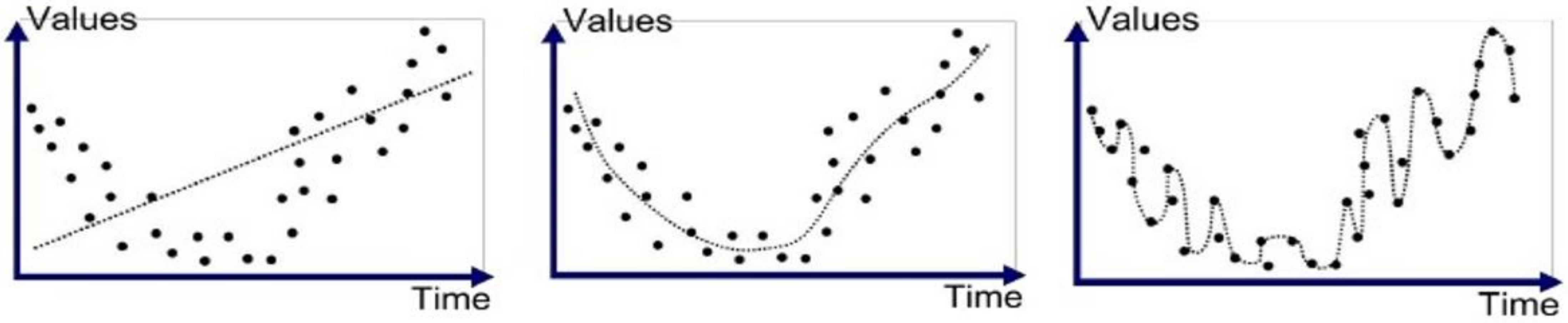
Time Series forecasting is an important area of forecasting in which past observation of the same variable are collected and analyzed to develop a model describing the underlying relationship. The model is then used to extrapolate the time series into the future. The ARIMA's major limitation is the pre-assumed linear form of the model. That is, linear correlation structure is assumed among the time series values and therefore, no non-linear patterns can be captured by the ARIMA model.

Artificial Neural Networks (ANNs)

ANNs have attracted increasing attentions in the domain of time series forecasting. One excellent feature of ANN, when applied to the time series forecasting problems is its capability to capture non-linear trend in modeling. One drawback of ANN is on its overfitting feature that happens when a neural network over learns during the training period.

Overfitting

Overfitting occurs when a statistical model or machine learning algorithm captures the noise of the data. Intuitively, overfitting occurs when the model or the algorithm fits the data too well. Specifically, overfitting occurs if the model or algorithm shows low bias but high variance. Overfitting is often a result of an excessively complicated model, and it can be prevented by fitting multiple models and using validation or cross-validation to compare their predictive accuracies on test data.



Underfitted Good Fit/Robust Overfitted

Figure 1: Graphs of underfitting (high bias, low variance) vs. overfitting (low bias , high variance)

Multilayered Perceptron Neural Network

- Is typically composed of several layers of nodes / several hidden layers;
- MLP training is a supervised training in that the desired response of the network or target value; and
- The training algorithm is used to find the weights that minimizes some overall error measure such as the sum square error.

Bayesian Regularization Neural Network

- Extra term E_W , added by BRNN to the objective function of early stopping given in equation which penalize large weights in anticipation of reaching a better generalization and smoother mapping;
- In BRNN, the networks are trained using supervised learning;
- The optimal regularization parameters α and β can be determine by the Bayesian Technique; and
- Uses the concept of Bayesian Inference in such a way that the weight distribution is made optimal to learn the correct function that relevantly maps the input to the output and it ensures that the network is not overfitting.

Hybrid Model

In recent times, many researchers who are finding difficulty in generating predictions have explored into creation of hybrid models. At least two models are combined to construct a more powerful prediction model that covers the shortcomings of the individual models when used alone.

Hybrid methodology that has both linear and non-linear modeling capabilities can be a good strategy for practical used. Thus, ARIMA model can cater the linear component then the residuals from the linear model will utilized only in non-linear modeling or artificial neural network.

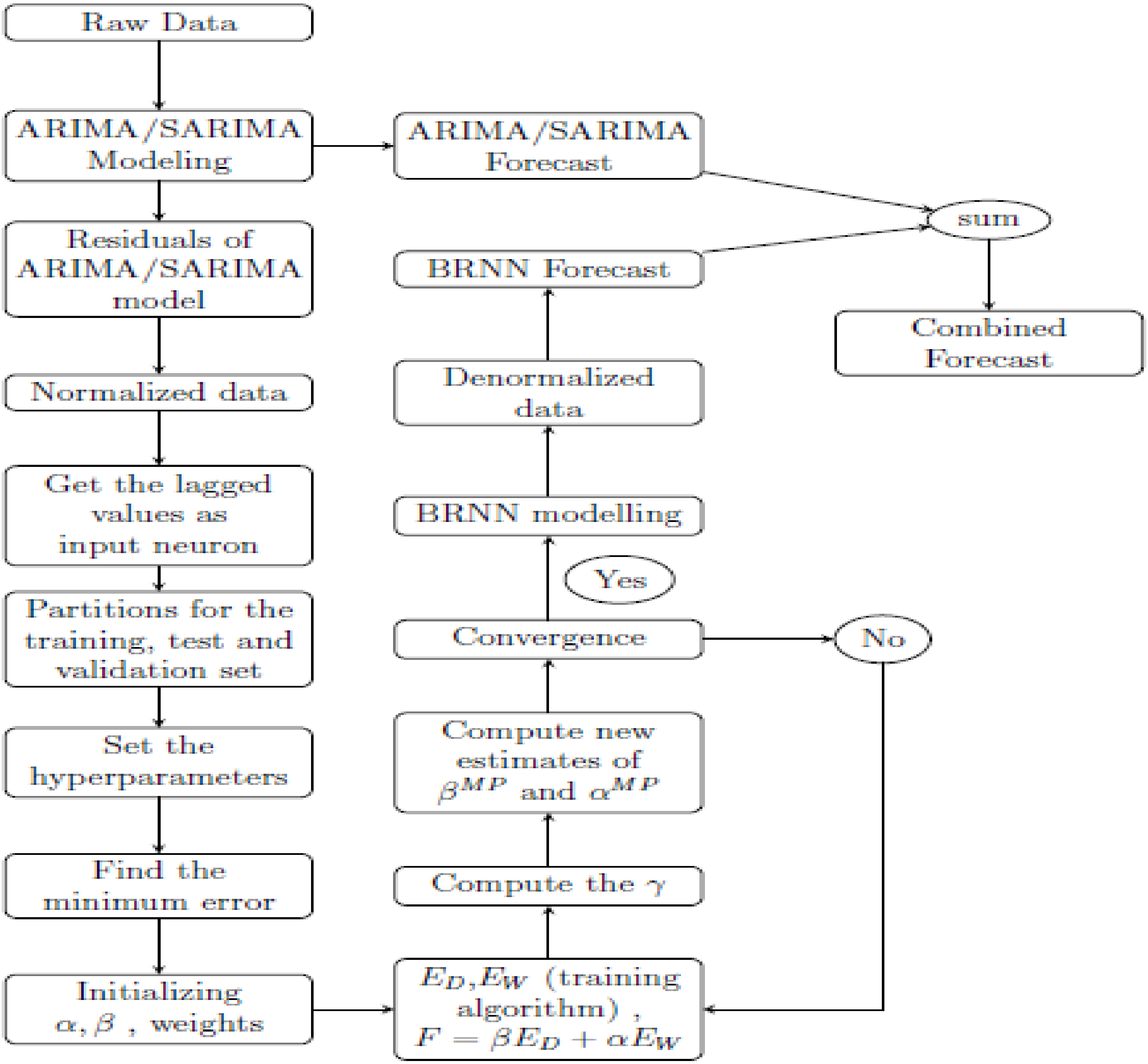


Figure 2: Hybrid ARIMA-BRNN Model and SARIMA-BRNN Model

Real Data

The time series of Gross Domestic Product of the Philippines (GDP) from year 1981 1st Quarter to year 2018 2nd Quarter, the residual of the model will be utilized for the hybridization process.

GDP series is composed of 150 observations. Therefore, the final model of the Gross Domestic Product of the Philippines (1981Q1-2018Q2) is

$$ARIMA(1, 1, 1) \times (1, 1, 1)_4$$

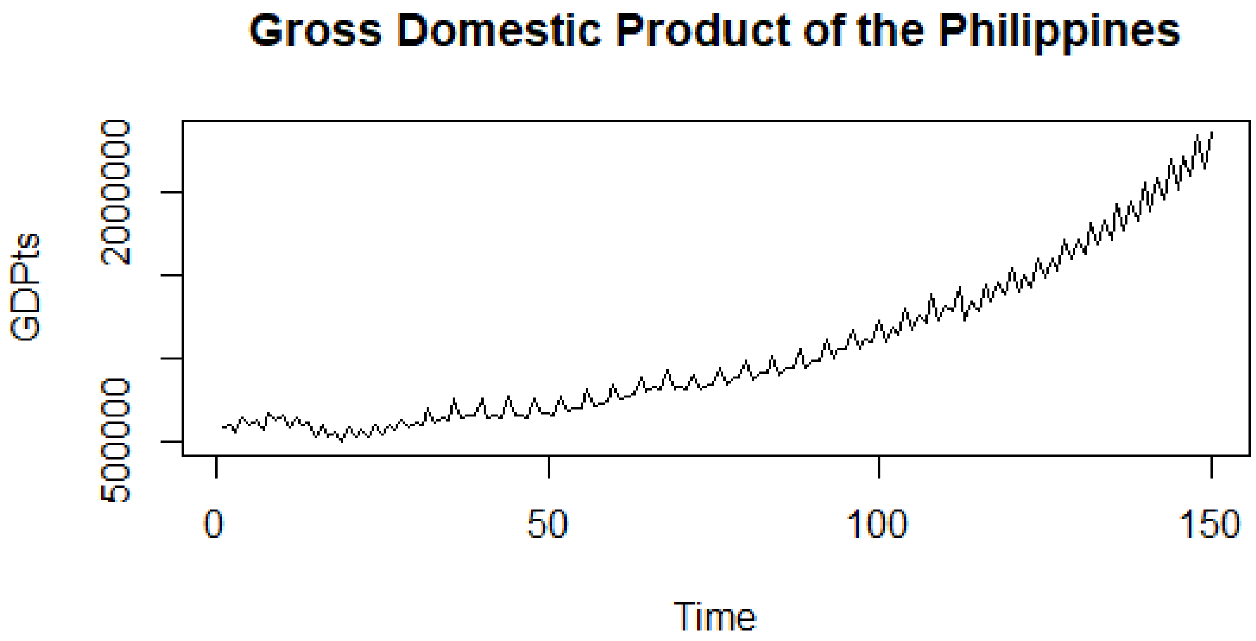


Figure 3: Time Series plot of GDP

Table 1: Data Partitioning of GDP Data

Partition	Number of Observation	Time Index
Training Set	116	5-120
Testing Set	20	121-140
Validation Set	10	141-150

Table 2: Setting the Hyperparameters

Number of Hidden neurons	**2, 3, 5, 9, 15**
Learning rate	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

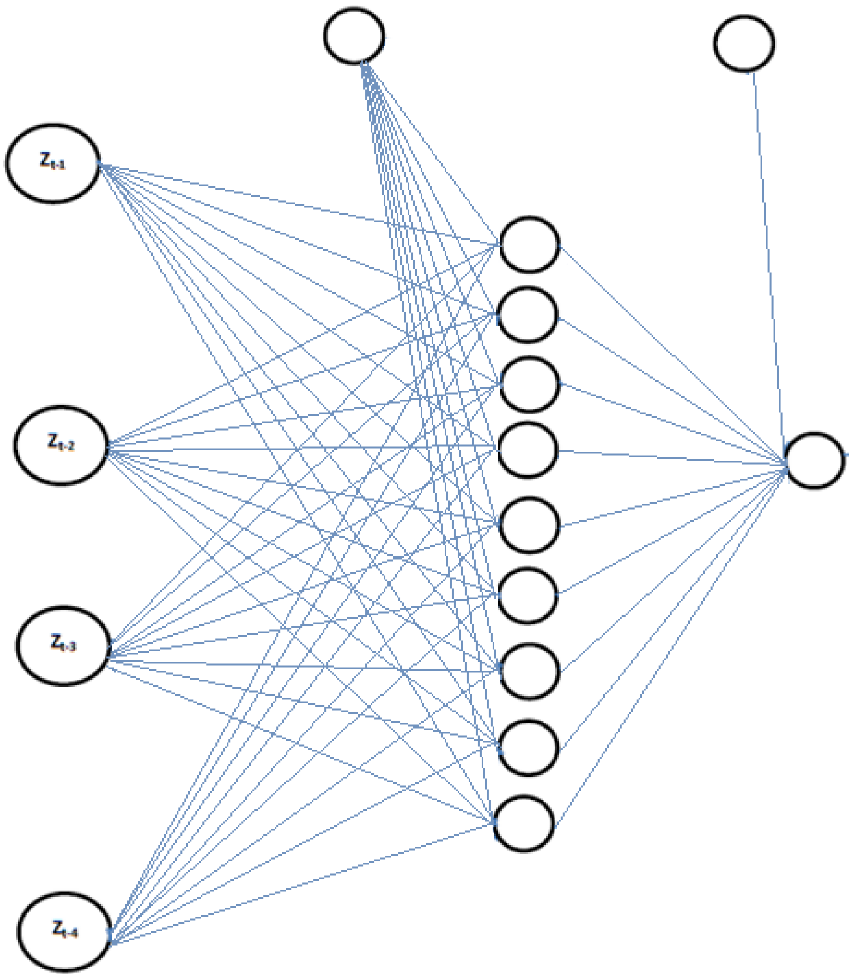


Figure 4: MLP (4-9-1) model architecture of GDP data

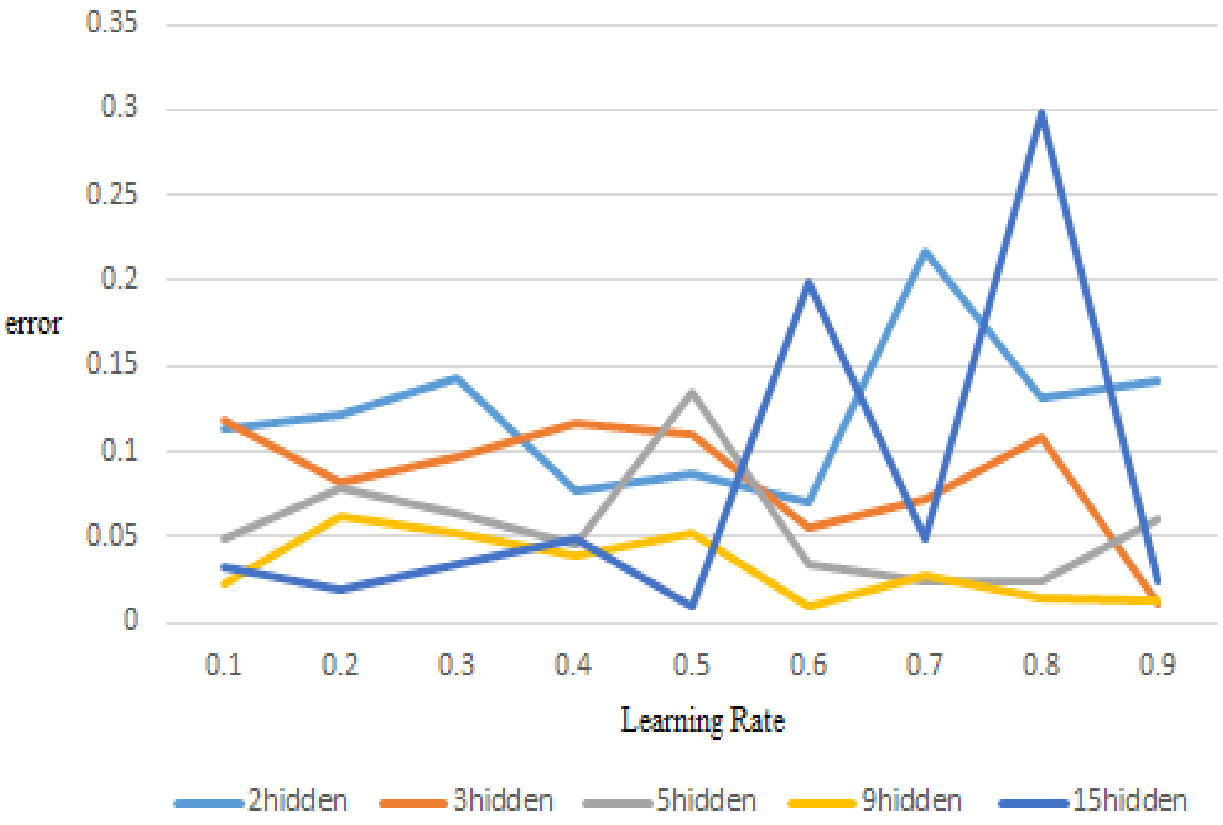


Figure 5: The minimum error located at learning rate of 0.6

Table 3: Forecasting performance for *i*-step ahead of ARIMA(1,1,1) x $(1,1,1)_4$, BRNN, MLP, SARIMA-BRNN and SARIMA-MLP where $i \in \{1,5,10\}$.

Models	1-Step Ahead		5-Step Ahead		10-Step Ahead	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
SARIMA	**1188.12**	**0.06298**	10527.01	0.452	100019.1	3.1707
BRNN	28516.88	1.5115	77017.21	3.254	60370.3	2.58934
SARIMA-BRNN	2411.124	0.1278	9520.368	0.417	**26085.7**	**0.9941**
MLP	80507.88	4.26735	143243.7	6.575	192372.7	8.2795
SARIMA-MLP	1931.021	0.10235	**5164.48**	**0.2038**	95383.67	2.897

In determining the BRNN model, recall that there are 4 lagged values as input neurons, 2 hidden neurons as the result for the best hyperparameters and one output neuron in the output layer. Hence, final model is BRNN(4-2-1). After getting the final model, we predict in 10-step ahead and denormalized the data for the forecasting accuracy.

Therefore, the Gross Domestic Product (GDP) data, SARIMA has a good forecasting performance in short term horizon while the SARIMA-BRNN has a good forecasting performance in long term horizon in terms of RMSE and MAPE. The result for the test of significant difference, SARIMA-BRNN is superior forecast accuracy compared to SARIMA and SARIMA-MLP for long term horizon.

Table 5: Test for significant difference between SARIMA vs. SARIMA-BRNN, SARIMA-BRNN vs. SARIMA-MLP and SARIMA vs. SARIMA-MLP.

Group Comparison	10-Step Ahead
SARIMA vs. SARIMA-BRNN	$4.33e^{-05**}$
SARIMA-BRNN vs. SARIM-MLP	0.001953^{**}
SARIMA vs. SARIMA-MLP	0.7394

Note *p*-value* and *p*-value** denotes a significantly different with *p*-value < 0.05 and *p*-value <0.01 respectively.

Acknowledge

This study has been supported by the Commission on Higher Education and Premiere Research Institute of Science and Mathematics (PRISM) of Mindanao State University – Iligan Institute of Technology (MSU-IIT), under the Applied Mathematics and Statistics Research Group for the lavish support by providing sufficient funds and assistance to present and successfully hurdle this paper.

References

- [1] Bailer-Jones, C., Gupta, R. and Singh, H., *An Introduction to Artificial Neural Networks,* Automated Data Analysis in Astronomy, Narosa Publishing House, New Delhi, India, (2001).
- [2] Foresee, F. D. and Hagan, M. T., *Gauss-Newton approximation to Bayesian Learning,* Proceedings of the 1997 International Joint Conference on Neural Networks, (1997).
- [3] MacKay, D., *A Practical Bayesian Framework for Backprop Networks*, California Institute of Technology 139-74, (1991).
- [4] Zhang, G. P., *Time series forecasting using a Hybrid ARIMA and Neural Network Model,* Georgia State University, University Plaza Altanta, USA, (2003).