Current Opinion
CO 4 (1), 426– 423 (2024)
Received 25 October 2023 | Revised 25 November 2023 | Accepted 25 December 2023 | Online Available 10 January 2024
https://doi.org/10.52845/currentopinion.v4i1.265

Current Opinion
Current Opinion

OPEN ACCESS JOURNAL **ISSN(O) 2795- 935X**

Original Article

Smoothing for Optimizing Arima and ANN Models for Predicting the Number of Treatments for TB Case Patients

Sudaryanto[1], Slamet Sudaryanto N*[2]

[1]Information Systems, Faculty of Computer Science, Dian Nuwantoro University Semarang

[2]Informatics Engineering, Faculty of Computer Science, Dian Nuwantoro University Semarang

Corresponding Author: Slamet Sudaryanto N

Abstract: Statistical models can be used to characterize numerical data to understand its behavior and patterns. For example, The TB care model can signal to local governments regarding when they should carry out prevention and prepare health care facilities. To find a model that can be optimized to predict the number of TB patient care occupancies so that the model has adequate performance. This research examines the performance of Wavelet-ARIMA-GARCH using tuning parameters in modeling and forecasting. Using Semarang City TB Incidence Treatment data from 2019 to 2022, this research concludes that the hybrid ANN -Wavelet-GARCH(1,1) model with parameter tuning is the best performance model.

Keywords : TB Patient Care, ARIMA, ANN, GARCH, Tunning Parameter

Introduction

Tuberculosis (TB) is still a major global health problem. It is estimated that the number of people diagnosed with TB in 2021 globally will be 10.6 million cases, an increase of around 600,000 cases from 2020, which was estimated at 10 million TB cases. Of the 10.6 million cases, there are 6.4 million (60.3%) people who have been reported and are undergoing treatment, and 4.2 million (39.7%) others who have not been found and reported. Based on the 2022 WHO report, every day almost 4,400 people lose their lives due to TB and almost 30,000 people fall ill due to TB transmission. (WHO, 2020).

In 2021, Indonesia was in third place with the highest number of cases, but in 2022 Indonesia rose to second place with the highest number of TB cases in the world after India, followed by China. Cases In 2022, the incidence of TB in Indonesia is estimated at 969,000 TB cases (one person every 33 seconds), this figure is up 17% from 2021, namely 824,000 cases. The prevalence of TB in Indonesia is 354 per 100,000 population, which means that for every 100,000 people in Indonesia, 354 people suffer from TB. This situation is a major obstacle to realizing the target of eliminating TB by 2030. To achieve this goal, TB prevention and control services must be provided in the context of broad health coverage with the support of information technology such as building TB incidence prediction models along with analysis of risk factors (Saifullah et al., 2021).

The number of TB incident treatments that occur is very fluctuating, so it is necessary to develop an accurate TB incidence treatment prediction model (Wang *et al.*, 2017). To obtain an accurate forecasting method, several hybrid approaches are used to predict fluctuating time-series data patterns (noise). To optimize forecasting methods on non-stationary data, signal decomposition as denoising is required. Optimization using signal decomposition to extract the part of the time-series signal that is not stationary can improve the performance of the prediction model (Cao *et al.*, 2013).

The use of wavelets for forecasting time series data, especially for fluctuating data, is experiencing rapid development. The wavelet transformation that is considered more suitable for time series data is the Discrete Wavelet Transform (DWT) because, at each level of decomposition, there are wavelet coefficients and scales as large as the length of the data. This advantage reduces the weakness of filtering with DWT (Discrete Wavelet Transform) which can be carried out on any sample size. Determination of decomposition levels and coefficients used as model input using multi-scale decomposition. The development carried out in this paper is a refinement of computational techniques so that the decomposition level and the number of coefficients at each level can be selected automatically based on predicted values that minimize error.

Deep Learning is a special subdivision of Machine Learning. It typically uses computational procedures known as neural networks (ANN) to achieve diverse goals. Artificial Neural Networks (ANN) are sequences of operations applied to data, resulting in the creation of filters capable of representing very complex functions. Generally, a neural network can be interpreted as a feature extractor followed by exploitation of the extracted features. Neural networks can cover a large number of trainable parameters (sometimes reaching tens of millions), which is usually achieved using the same methodology used in Machine Learning as a whole. However, it is rare to use trust region algorithms, with stochastic gradient descent-based methods such as Adam (Kingma and Ba, 2014) and rmsProp (Zhou *et al.*, 2018) become the preferred alternative.

Determining the appropriate weights in the network poses a challenge for Artificial Neural Networks (ANN). A comparative study was conducted to evaluate the efficacy of Artificial Neural Network (ANN) based training in conjunction with Gradient Descent and Genetic Algorithm (GA). The results show that GA shows a slight advantage in terms of Mean Square Error (MSE) when applied to cancer datasets for average classification error. In contrast, Gradient Descent shows superior performance when applied to the diabetes dataset. It is evident from this study that further experiments with additional data sets are needed to improve the efficacy of ANN training (Hassanien et al., 2018).

To improve gradient descent techniques in artificial neural networks, such as quickprop, backpropagation, Delta-Bar-Delta, and Super SAB, the function approximation error is evaluated using quadratic polynomials to achieve a minimum quadratic error function. The modified approach of the partial derivative method in the weight update process in the backpropagation algorithm allows for adjusting the learning rate for each weight in the neural network. The improved gradient descent method outperforms standard gradient descent and momentum gradient descent techniques (Popa, 2015).

This research paper will improve the optimization of forecasting methods by carrying out decomposition as noise smoothing. First stage decomposition using the wavelet thresholding method. Then for the second stage use ARIMA, ANN, and GARCH models to estimate fluctuations in the number of TB incidences. The significance of the wavelet decomposition boundary problem is explained through the application of two different decomposition approaches. Next, the impact of the detail component on forecasting is evaluated by comparing the forecast results with and without the inclusion of the detail component. ARIMA and ANN models are used to predict the approximation component, while ANN and GARCH models are used to predict the detail component.

Literature Review

Wavelet Analysis

Wavelet analysis has emerged as a prominent mathematical technique for signal and image analysis. In recent years, the wavelet transform has gained wide popularity due to its ability to effectively describe nonstationary processes. Compared with the Fourier transform, the use of wavelets is much wider and attracts great attention. This is mainly attributed to the extraordinary ability of wavelets to analyze various types of data, including stationary and nonstationary data, as well as estimate smooth functions. In contrast, the Fourier transform shows certain limitations as an analytical tool, especially when dealing with nonstationary data. These limitations include the inability to localize the time domain and the relatively greater computational complexity associated with decomposition algorithms.

The study of noise reduction has been a pervasive aspect of signal estimation across a variety of scientific disciplines over a considerable period. It has been revealed through recent research that noise acts as a barrier to the effectiveness of various methodologies, including identification, parameter estimation, and prediction accuracy (Liu *et al.*, 2019). As a result, there is a strong preference for early manipulation of the data to minimize noise interference without compromising the inherent dynamics of the underlying signal (Bing *et al.*, 2020).

Discrete Wavelet Decomposition (DWD).

DWD is a pre-processing method that facilitates the projection of time series onto a set of orthonormal basis functions. This particular transformation is implemented to extract additional insights from the original time domain data. After applying DWD to the data, signal analysis can be performed by decoding it at various frequencies. While high-frequency components can introduce noise, low-frequency components tend to show visible patterns derived from the original data, thereby facilitating the estimation process. In this particular investigation, DWD is used to decompose weekly Henry Hub spot prices into four distinct subseries.

Heteroscedasticity

The error factor in a regression model usually has problems with violations of assumptions in the residuals. A situation is said to be heteroscedasticity if the data has an error variance that is not constant for each observation or in other words, violates the $Var\,\varepsilon_t = \sigma^2$ assumption (Rosopa, Schaffer and Schroeder, 2013)

Generalized Autoregressive Conditional Heteroscesdasticity (GARCH)

Studies from 2002 to 2011 associated with higher volatility in that period. (Garcia *et al.*, 2005). Therefore, it is deemed appropriate to investigate the performance of hybrid ARIMA with a volatility model because the ARIMA model alone is not able to handle the volatility present in the data series. Previous research shows that the generalized autoregressive conditional heteroscedasticity (GARCH) model is widely applied to handle gold price volatility. A hybrid model combining the strengths of ARIMA-GARCH represents a promising approach in modeling and forecasting daily gold prices (Sun and Yu, 2020).

Research Methods

In this particular research, research was carried out to investigate the utilization of wavelets in approximating non-linear functions. The purpose of this stage is first-stage noise reduction. The decomposition functions are evaluated using different wavelet bases. Each basis wavelet is subject to a combination of resolution level and threshold function. The newly obtained wavelet coefficients are then reconstructed using the inverse wavelet transform, thereby returning the signal to its original form. This is accompanied by the generation of new signal estimation results through the application of a threshold wavelet approach. Then calculate the Cross-Validation (CV) value of the model used in the approximation process. The CV value is used as an indicator of the goodness of the model to obtain an optimal model in the approximation process using the wavelet-thresholding method. The computational aspects of this research were carried out using the Python programming language. In terms of data, a total of 148 data sequences in time series format were used to simulate each function.

After denoising using the Wavelet Threshold approach, the noise condition has been reduced but still has several error components as heteroscedasticity. This means that the data still has noise which requires a second stage of denoising using the GARH approach. GARCH is a model used in forecasting data that has heteroscedasticity problems.

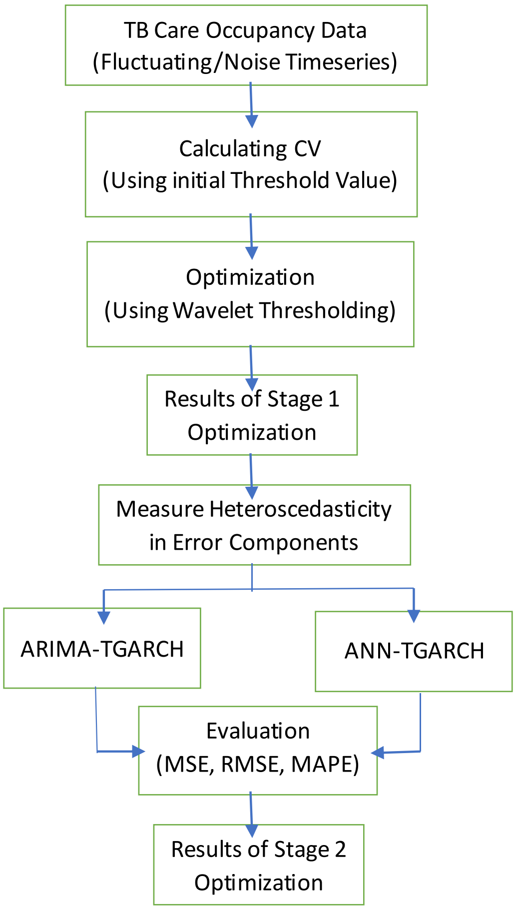


Figure 1. Overall system model of the present study

The level of smoothness of the curve in the Function Approach using the wavelet thresholding method is influenced by various parameters. These parameters include the selected wavelet function type, resolution level, thresholding function type, and thresholding value (IEEE Communications Society and Institute of Electrical and Electronics Engineers, no date). To obtain optimal results, optimization is carried out for each parameter. The optimization process consists of two parts. The first part focuses on optimizing the wavelet basis and resolution levels used in the wavelet transform process, as well as the threshold function used in the thresholding process (Bayer, Kozakevicius and Cintra, 2019). Next, the second part involves optimizing specific threshold values that serve as threshold limits in the thresholding process. This particular optimization is achieved through the utilization of the GARCH optimization method.

Result and Discussion

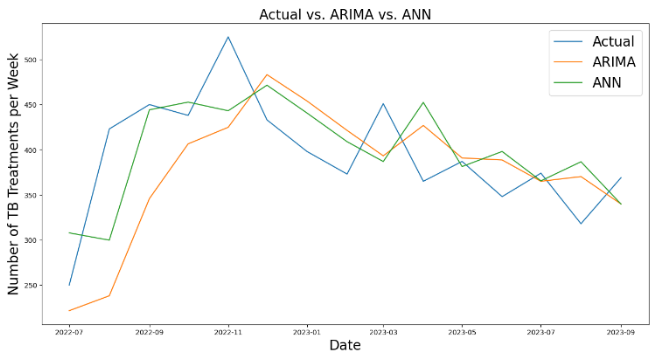
Wavelet Decomposition Results

Figure 2. Shows the components resulting from the discrete wavelet decomposition process at level 3. Starting from the bottom up, we observe the approximation components and detailed components with the approach decomposed up to level three. The approximation and detail components are presented in three sequences. This discrete wavelet decomposition can be achieved through various approaches such as Daubechies, Coiflets, Symlets, or Discrete Meyer. Among these wavelets, Daubechies and Symlet allow perfect reconstruction with the maximum number of missing moments. Symlets show perfect symmetry, whereas Daubechies do not. Due to the fact that symmetry can limit flexibility in representing data, Daubechies wavelets were chosen. Three optimal mean square errors (MSE) are used to determine the total number of missing moments of the Daubechies wavelet within the available perfect reconstruction range.

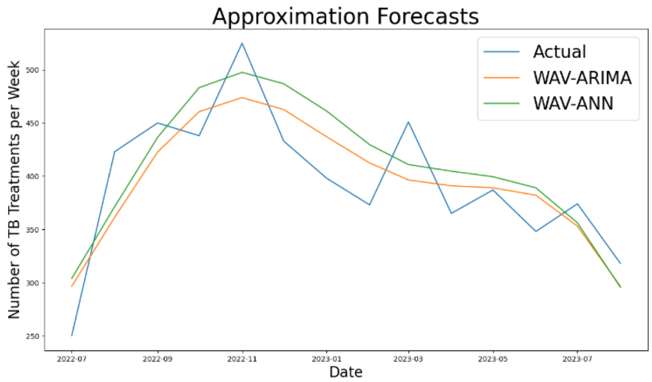
Figure 2. Level 3 Component Decomposition

ARIMA Prediction Results, ANN Combination with Wavelet

In the first group of predictions, ANN and ARIMA are used to predict without noise intervention on time-series data, or ARIMA and ANN are applied, and the original data is predicted directly. Next, the results of the decomposition of time-series data with wavelets are combined with ARIMA and then with ANN to predict TB incidence care occupancy using a multi-step process in the future. In the case of combined wavelet decomposition, only the approximate components are used for forecasting. The optimal time lag for each case is selected based on the AR term in the ARIMA model. For clarity, only two models are compared in each figure, namely ARIMA and ANN.



Figur 3. ARIMA Vs. Prediction ANN based on time series data (without noise)



Figur 4. Wavelet-ARIMA Prediction Vs. Wavelet- ANN based on timeseries data that has been decomposed to level 3.

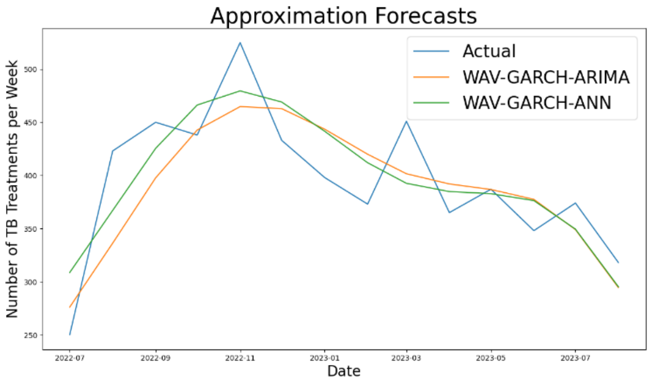


Figure 5. Wavelet-GARCH-ARIMA Prediction Vs. Wavelet-GARCH-ANN is based on timeseries data that has been decomposed to level 3.

Figure 3. Above is a prediction from the ARIMA and ANN models using time-series data without denoising first. Figure 4. Represents the prediction results from a combination of the RIMA and ANN models using time-series data that has been decomposed using wavelets. The figure shows that ANN and ARIMA are susceptible to data fluctuations, while the combination scenario with wavelet decomposition appears to produce a smoother wavelet decomposition. Primary data is decomposed into subseries via wavelet decomposition, resulting in data that is more suitable for prediction. So this process improves prediction performance.

ARIMA Prediction Results, ANN Combination with Wavelet and GARCH

The estimated components used in this section are estimated using ARIMA and ANN methods, while the detailed components are combined with GARCH. By comparing subsequent models, we can ascertain which model is more appropriate for forecasting detailed components. Furthermore, GARCH is commonly used in circuits that exhibit high fluctuations caused by unpredictable random effects (Garcia et al., 2005). Given that detailed components are highly volatile signals characterized by heteroscedasticity, we use ARIMA, ANN, and GARCH techniques to estimate them. Figure 5 is a finding that reveals that the detailed component has minimal impact on ARIMA forecasting results, considering that it reduces forecasting performance when ANN is used as the forecasting method. Alternative approaches are also applied to estimate detailed components. Engle's ARCH test identified the presence of heteroscedasticity in the detailed components. However, the utilization of GARCH leads to a slight improvement in forecasting performance. As a result, we obtain two conclusions. First, GARCH is more suitable for forecasting components in detail. Second, although the granular component exerts a small influence, it can indeed be negligible or even detrimental. Thus, it becomes difficult to assert that there is any practical benefit to including detailed components in the model.

Table 1. Comparison of performance results against ARIMA and ANN

Metric	MSE	RMSE	MAPE
ARIMA	5155.84	71.80	0.15
ANN	3331.51	57.72	0.12
Wavelet with ARIMA	5208.00	72.15	0.14
Wavelet with ANN	3069.48	55.40	0.10
Wavelet with GARCH and ARIMA	2869.66	48.90	0.98
Wavelet with with GARCH and ANN	2869.07	37.21	0.26

Conclusion

Comparison between the results presented in the evaluation matrix in Table 1 shows that the wavelet decomposition level 3 process can provide performance improvements for both ARIMA and ANN models. This can be seen from the MAPE evaluation results where before wavelet decomposition was carried out there was a decrease in presentation for ARIMA of 0.3 and 0.1 for ANN. Meanwhile, the combination of wavelet and GARCH reduces the MSE and RMSE values slightly but increases the MAPE value slightly. Such results, if seen from the MSE and RMSE values which are still quite large, indicate that the model is still vulnerable or sensitive to extreme values (outliers) because it calculates errors squarely. Outliers can make a significant contribution to the MSE and RMSE values. A still high value for MSE/RMSE indicates that the absolute prediction error is greater. This may reflect a significant degree of variation or deterioration in the predicted values. For this reason, to improve the optimization of ARIMA and ANN models, methods are still needed to manage outliers in the threshold management model, both for wavelets and GARCH thresholds. Our presented estimation results show that GARCH is a more suitable ANN than ARIMA for predicting detailed components. However, incorporating detailed components into the forecasting model does not produce significant advantages over models that only use the detailed components of the forecast. This conclusion is supported by tuning parameters where the optimum parameters used are num_lags = 3, hiden_layer_sizes = 4, and decomposition level = 3. Meanwhile, the ARIMA order used is (2,1,1).

References

1. Aderita, N.I., Murti, B. and Suryani, N. (2016) 'Risk Factors Affecting Multi-Drug Resistant Tuberculosis in Surakarta and Wonogiri, Central Java, Indonesia', *Journal of Epidemiology and PublicHealth*, 01(02), pp. 86–99. Available at: https://doi.org/10.26911/jepublichealth.2016.01.02.02.
2. Bayer, F.M., Kozakevicius, A.J. and Cintra, R.J. (2019) 'An iterative wavelet threshold for signal denoising', *Signal Processing*, 162, pp. 10–20. Available at: https://doi.org/10.1016/j.sigpro.2019.04.005.
3. Bing, P. *et al.* (2020) 'Noise Reduction in ECG Signal Using an Effective Hybrid Scheme', *IEEE Access*, 8, pp. 160790–160801. Available at: https://doi.org/10.1109/ACCESS.2020.3021068.
4. Cao, S. *et al.* (2013) *A hybrid seasonal prediction model for tuberculosis incidence in China*. Available at: http://www.biomedcentral.com/1472-6947/13/56.
5. Garcia, R.C. *et al.* (2005) 'A GARCH forecasting model to predict day-ahead electricity prices', *IEEE Transactions on Power Systems*, 20(2), pp. 867–874. Available at: https://doi.org/10.1109/TPWRS.2005.846044.
6. Hassanien, A.E. *et al.* (no date) *10th International Conference on Intelligent Systems Design and Applications (ISDA), 2010 Nov. 29, 2010 - Dec. 1, 2010, Cairo, Egypt ; [including workshop papers]*.
7. IEEE Communications Society and Institute of Electrical and Electronics Engineers (no date) *SETIT 2012 : Sciences of Electronics, Technologies of Information and Telecommunications, Sousse, 21-24 March 2012*, *Tunisia*.
8. Kingma, D.P. and Ba, J. (2014) 'Adam: A Method for Stochastic Optimization'. Available at:http://arxiv.org/abs/1412.6980
9. Liu, F. *et al.* (2019) 'An Ensemble Model Based on Adaptive Noise Reducer and Over-Fitting Prevention LSTM for Multivariate Time Series Forecasting', *IEEE Access*, 7, pp. 26102–26115. Available at: https://doi.org/10.1109/ACC ESS.2019.2900371.
10. Popa, C.A. (2015) 'Enhanced gradient descent algorithms for complex-valued neural networks', in *Proceedings - 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2014*. Institute of Electrical and Electronics Engineers Inc., pp.272–279. Available at: https://doi.org/10.1109/SYNASC.2014.44.
11. Rosopa, P.J., Schaffer, M.M. and Schroeder , A.N. (2013) 'Managing heteroscedasticity in general linear models' , *Psychological Methods*, 18(3), pp. 335–351. Available at: https://doi.org/10.1037/ a0032553.
12. Saifullah, A. *et al.* (2021) 'Evaluation of risk factors associated with the development of MDR-and XDR-TB in a tertiary care hospital: A retrospective cohort study', *PeerJ*, 9, pp. 1–19. Available at: https://doi.org/10.7717/peerj.10826.

13. Sun, H. and Yu, B. (2020) 'Volatility asymmetry in functional threshold GARCH model', *Journal of Time Series Analysis*, 41(1), pp. 95–109. Available at: https://doi.org/10.1111/jtsa.12495.
14. Wang, K.W. *et al.* (2017) 'Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network', *Epidemiology and Infection*, 145(6), pp. 1118–1129. Available at: https://doi.org/10.1017/S0950268816003216.
15. WHO (2020) *Consolidated Guidelines on Tuberculosis Treatment*, *Who*.
16. Yang, Z. *et al.* (2004) *Identification of Risk Factors for Extrapulmonary Tuberculosis*, *Extrapulmonary Tuberculosis Risk Factors • CID*. Available at: https://academic.oup.com/cid/article/38/2/199/287082.
17. Zhou, Y. *et al.* (2018) 'Location Nearest Neighbor Query Method Based on Voronoi Map', *Beijing Gongye Daxue Xuebao/Journal of Beijing University of Technology* [Preprint]. Available at: https://doi.org/10.11936/bjutxb2017040016