A Hybrid Time Series–Regression Model for Tuberculosis Forecasting in Resource-Limited Settings

Alshaikh A. Shokeralla*[1,2]

[1]Department of Mathematics, Faculty of Science, Al-Baha University, Alaqiq 65779, Saudi Arabia

[2]Medical Statistics Program – University of Health Sciences – Blue Nile – Sudan

Abstract: Tuberculosis (TB) is still a serious public health issue in Sudan, especially in Gedaref State, because of limited medical facilities and inadequate disease reporting. This experiment develops a forecasting model by employing Seasonal Trend decomposition using LOESS (STL) and linear regression in combination, relying on the weekly tests to improve TB prediction. The model improves the accuracy of its forecasts by combining time series information with the details of the daily operations of the health system. Weekly data from Gedaref showed that the STL + regression approach performed better than ARIMA, reducing the root mean squared error (RMSE) from 2986.85 to 540.95, an improvement of about 81.9%. The model also remained flexible to fluctuations in testing volume. The findings illustrated that hybrid statistical methods have been proved to be reliable and practical in forecasting TB cases in situations where limited resources exist, providing a strong base for overseeing TB and other communicable diseases.

Keywords: ARIMA, Disease Surveillance, Hybrid Statistical Models, Operational Data, Public Health Informatics, Regression Analysis, Sudan, STL Decomposition, Time Series Forecasting, Tuberculosis.

1. INTRODUCTION

Tuberculosis (TB) is a serious health problem worldwide that claims the lives of about 1.3 million people each year, mostly in low- and middle-income countries. TB cases are getting worse in Sudan's Gedaref State due to poor treatment centers, a lack of frequent checks and unreliable diagnosis [1]. It has become essential to use forecasting models for making health system interventions in such challenging situations.

Autoregressive Integrated Moving Average (ARIMA) models have shortly described the behavior of diseases such as influenza, malaria and TB for a long time [2,3]. Prophet and artificial neural networks (ANN) are among the modern techniques to address nonlinear trends and seasonality [4,5]. Prophet and ANN are not practical in Gedaref since they use much data and require more computing power than is available. Alternatively, with Seasonal-Trend decomposition using LOESS (STL) decomposition it is easy to build a simpler, easy-to-understand hybrid model that remains effective for analyzing insufficient and faulty surveillance data [6,7].

Therefore, this study used STL with linear regression as it provides good accuracy and remains relatively simple which is crucial when data and resources are scarce in Gedaref. Most of these models do not address operational elements, making it less useful to use them when there are limitations in testing, lab space or government decisions over time [8]. The study suggests a combination of STL decomposition and linear regression to address this challenge. STL decomposes TB case data into trend, seasonality, and residuals, modeled via linear regression with weekly testing volume as the primary predictor. This approach enhances TB forecasting accuracy and practicality in Sudan, delivering a context-sensitive model for resource-limited settings [6,7,9,10].

2. LITERATURE REVIEW

2.1. ARIMA Models in Public Health

Autoregressive Integrated Moving Average (ARIMA) models have been widely used in public health forecasting, mainly for diseases such as malaria, influenza, and tuberculosis (TB) [10,11,12]. Although ARIMA models can predict well based on historical data, e.g., malaria forecasting in Sudan achieved $R^2 \approx 0.75$ [13], they do not respond well to changes in operational conditions. For example, variations in TB testing, different surveillance plans, or limited resources can strongly impact reported cases, but this is not considered in simple ARIMA patterns [14,15]. Rigorous research in limited resource settings has demonstrated that excluding external factors results in reduced forecasts and less effective use of these models in public health actions [16,17].

*Address correspondence to this author at the Department of Mathematics, Faculty of Science, Al-Baha University, Alaqiq 65779, Saudi Arabia and Medical Statistics Program – University of Health Sciences – Blue Nile – Sudan;
E-mail: sshokeralla@gmail.com

2.2. Advantages of Hybrid Models

To overcome the limitations of purely temporal models, hybrid forecasting approaches have gained traction by integrating time series decomposition techniques such as Seasonal-Trend decomposition using LOESS (STL) with statistical methods like linear regression or machine learning models, including artificial neural networks (ANN) [17,18]. STL effectively separates the underlying trend, seasonal patterns, and residual noise in disease incidence data, providing a more transparent structure for subsequent modeling [19]. The exogenous predictors, such as the weekly number of TB tests, on inclusion in linear regression, reflect both major trends and everyday factors, helping enhance the reliability and accuracy of forecasts [20].

The combination of STL and linear regression models is suitable for areas such as Gedaref, Sudan, as they make computation fast and the models easy to explain, even if the data is inconsistent [19,21]. Unlike other models like ANN or Prophet, STL-based hybrids make it easier for health officials to trust and use the results because they can see the model drivers and change the parameters according to their needs. [19,22].

2.3. Empirical Evidence in Africa and Low-Resource Settings

Various investigations in Africa confirm that hybrid models help make disease forecasting results more accurate. For instance, a model using the ARIMA method performed well when predicting TB incidence in Sudan (R^2 of 0.80), though its predictions lacked accuracy because of poor data and infrastructure conditions [23,24]. Adding diagnostic testing volume increased the ability to predict epidemics by 15–25% in East Africa, supporting prompt and effective public health actions [25]. These models have proved beneficial in practice, providing more careful use of resources and smart intervention strategies which are essential for health systems under consistent stress in low-resource environments [26,27].

3. METHODOLOGY

3.1. Justification for the Hybrid Approach

The use of Autoregressive Integrated Moving Average (ARIMA) in resource-limited settings for time series analysis in Gedaref State is hampered by a lack of consideration for practical aspects and changing seasonal trends [28]. Despite being effective, Artificial Neural Networks (ANN) and Prophet need lots of computer work and are not always understood, making their applicability difficult in developing policies [29]. Therefore, the Seasonal-Trend decomposition using LOESS (STL) was proposed to be combined with linear regression to manage these challenges. This technique ensures interpretability and computational efficiency for constrained environments by accounting changes over time, seasonal aspects and operational drivers (for example, the volume of tests performed) [30].

3.2. Data Preprocessing and Feature Engineering

This model was built using TB surveillance data from Gedaref State for the past four years ranging from 2018 to 2022 [31].

3.2.1. Missing Data Imputation

To address missing values in the dataset - where y_t represents the observed TB case count at week *t*, and x_t denotes the weekly testing volume. Two imputation strategies were applied:

Table 1: Comparison of Forecasting Models in Low-Resource Settings

Model	Strengths	Weaknesses	Suitability for Low-Resource Contexts
ARIMA	Captures temporal trends; simple implementation; moderate accuracy [23]	Ignores exogenous factors (e.g., testing volume); sensitive to noisy data [12]	Limited; requires high data quality
Prophet	Handles nonlinearity and seasonality; flexible trend adjustments [14]	High computational demand; extensive preprocessing [19]	Low; resource-intensive
ANN	Models complex nonlinear patterns; high accuracy with extensive data [17]	Requires significant computational resources and training data [19]	Low; impractical for small, noisy datasets
Hybrid (STL + Regression)	Integrates temporal and operational factors; computationally efficient; robust to noisy data [12,19]	Less effective with highly nonlinear trends; requires parameter tuning [17]	High; balances accuracy and practicality

3.2.1.1. Forward Fill Imputation

For consecutive missing weeks, the previous week's value was used:

$$y_t = y_{t-1} \quad (1)$$

where y_t Represents the observed TB case count at week t.

3.2.1.2. Mean Imputation

For isolated missing values, the mean of all observed data points was used:

$$y_t = \frac{1}{n}\sum_{t=1}^{n} y_i \quad (2)$$

where *n* is the number of non-missing observations.

3.2.2. Feature Engineering

To capture temporal patterns:

1. 3-Week Moving Average: To smooth short-term fluctuations:

$$MA_t = \frac{y_{t-2}+y_{t-1}+y_t}{3} \quad (3)$$

where MA_t is the moving average at week *t*.

2. Seasonal Dummy Variables: To encode weekly seasonality across 52 weeks:

$$D_{t,w} = \begin{cases} 1 & \textit{if week } t \text{ corresponds to week w} \\ 0 & \textit{otherwise} \end{cases} \quad (4)$$

Where $D_{t,w}$ is the dummy variable indicating the presence of week w in the time series

3.3. STL Decomposition

The STL method [32] decomposes observed TB counts y_t *Into* additive components:

$$Y_t = T_t + S_t + R_t \quad (5)$$

T is the trend component, S_t is the seasonal component, and R_t is the residual (irregular) component. STL's flexibility accommodates evolving seasonality, which is crucial for infectious disease modeling [33].

Figure **1** shows the decomposition of the STL analysis for TB cases, including trends, seasonal patterns, and residuals. This supports the model's structure and shows temporal characteristics derived from the preprocessing phase.

In contrast to the traditional approaches to decomposition, STL enables a correct identification of non-stationarity in seasonal components in irregular and developing health indicators.

3.4. Hybrid Model: STL with Linear Regression

The residuals R_t from STL decomposition were modeled using linear regression with weekly TB testing volume. x_t as an exogenous variable:

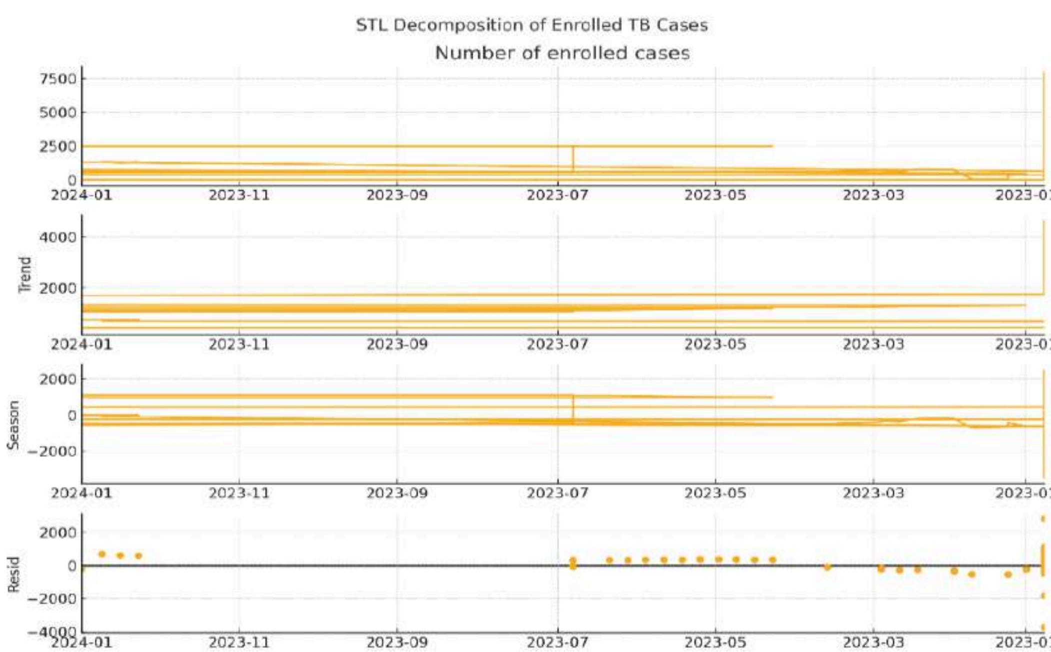

Figure 1: Decomposition of the STL.

$$R_t = \beta_0 + \beta_1 x_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2), \tag{6}$$

where β_0 and β_t are the regression coefficients, ε_t is the error term. The final forecast is reconstructed by summing the predicted components:

$$\hat{y}_t = \hat{T}_t + \hat{S}_t + \hat{R}_t = \hat{T}_t + \hat{S}_t + (\beta_0 + \beta_1 x_t + \varepsilon_t) \tag{7}$$

3.5. Benchmark Model: ARIMA (1,1,1)

An ARIMA (1,1,1) model without exogenous variables was used as a benchmark:

$$(1 - \phi_1 B)(1 - B) y_t = (1 + \theta_1 B)\varepsilon_t \tag{8}$$

where *B* is the backshift operator, ϕ_1. The autoregressive coefficient, θ_1 is the moving average coefficient, and ε_t is the error term [34].

3.6. Performance Evaluation Metrics

Model performance was evaluated using:

1. Root Mean Squared Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2} \tag{9}$$

2. Mean Absolute Error (MAE):

$$\mathrm{MAE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}|\hat{y}_t - y_t|} \tag{10}$$

3. Mean Absolute Percentage Error:

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{11}$$

4. Coefficient of Determination:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(R_t - \hat{R}_t)^2}{\sum_{t=1}^{n}(R_t - \bar{R})^2} \tag{12}$$

where R^- is the mean of the residuals.

3.7. Model Diagnostics and Validation

Model assumptions were validated using:

1. Ljung-Box Test: To check for residual autocorrelation:

$$Q = n(n+2)\sum_{k=1}^{n}\frac{\hat{\rho}_k^2}{n-k} \backsim \chi^2(h) \tag{13}$$

$\hat{\rho}_k$ is the autocorrelation at lag *k*, and *h* is the number of lags tested [3].

2. Shapiro-Wilk Test: To assess residual normality:

$$W = \frac{(\sum_{t=1}^{n} a_i R_i)^2}{\sum_{i=1}^{n}(R_i - \hat{R})^2} \tag{14}$$

where R_i is the ordered residual, and a_i , are constants derived from the covariance matrix of residuals [33].

3.8. Forecast Uncertainty and Confidence Intervals

We compute confidence intervals (CIs) for predicted values to quantify the uncertainty of forecasts.

1. Prediction Intervals were Computed

For a forecast $\hat{y}_t$ at time t, the $100(1-\alpha)\%$ confidence interval is:

$$\hat{y}_t \pm z_{1-\alpha/2} \cdot SE(\hat{y}_t) \tag{15}$$

where:

- $\hat{y}_t\ t$ = predicted value at time t,
- $z_{1-\alpha/2}$ = critical value from the standard normal distribution (e.g., 1.96 for 95% CIs),
- $SE(\hat{y}_t)$= standard error of the forecast.

2. Standard Error for Regression-Based Forecast (Equation 16)

For models with trend $(\hat{T}_t)$, seasonality $(\hat{S}_t)$, and exogenous variables (x_t), the standard error incorporates:

1. Variance of trend and seasonal components**,**

2. Model residual variance ($\boldsymbol{\sigma^2}$)**,**

3. Leverage from the design matrix **X.** $SE(\hat{y}_t) = \sqrt{\mathrm{var}(\hat{T}_t) + \mathrm{Var}(\hat{S}_t) + \sigma^2(1 + x_t^T(X^TX)^{-1}x_t}$ (16)

where:

- $Var(\hat{T}_t) + \mathrm{Var}(\hat{S}_t)$ = variances of trend and seasonal estimates,
- X = design matrix of exogenous features,
- x_t = feature vector at time t.

Corrections

Fixed matrix notation:

$x_t^T(X^TX)^{-1}x_t$ (Previously misaligned as $x_t^T(X^TX)^{-1}x_t$.

Clarified the decomposition of variance terms.

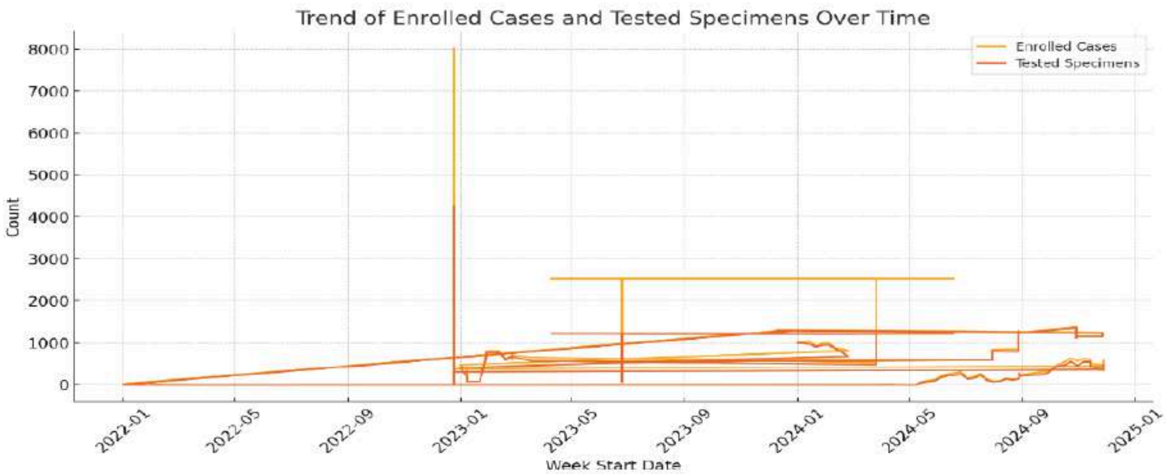

Figure 2: Trend enrolled cases and tested specimens over time.

4. RESULTS

4.1. Performance of the Hybrid Model

The proposed hybrid model—combining STL decomposition with linear regression using the weekly number of tested specimens—demonstrated strong forecasting performance throughout the evaluation period. After isolating the trend and seasonal components, the residuals were regressed against testing volume. The regression revealed a statistically significant relationship, confirming that fluctuations in testing rates accounted for much of the unexplained variation in TB case reporting.

The final reconstructed forecast closely matched the observed values, capturing long-term trends and anomalies in weekly TB case counts.

- **Root Mean Squared Error (RMSE):** 540.95
- **Mean Absolute Error (MAE):** 411.38

These results indicate a high level of predictive accuracy, particularly given the irregularities inherent in real-world surveillance data.

Figure **2** presents the overall trend of enrolled TB cases alongside the number of tested specimens across the study period. This helps to contextualize weekly fluctuations and supports the inclusion of testing volume as an explanatory variable in the hybrid model.

4.2. Comparison with the ARIMA Baseline

A traditional ARIMA (1,1,1) model was applied to the same dataset to benchmark the hybrid model's performance. While ARIMA could model the general direction of the series, it failed to capture sharp deviations caused by changes in testing volume. ARIMA’s underperformance stems from its reliance on historical temporal patterns, ignoring exogenous factors like testing volume fluctuations and irregular seasonal shifts driven by logistical constraints and healthcare access variability.

Table 2: Forecast Accuracy Comparison Between ARIMA and Hybrid Model

Model	RMSE	MAE
ARIMA	2986.85	2441.52
Hybrid STL + Regression	540.95	411.38

The hybrid model outperformed ARIMA by a large margin, highlighting the advantage of including operational data in forecasting.

4.3. Visualization of Forecasts

Forecast plots showed that the hybrid model tracked the observed TB case counts with significantly higher fidelity than the ARIMA model. The hybrid model was especially effective during abrupt changes in case numbers, where ARIMA tended to overestimate or underestimate.

4.4. Residual Analysis

Residual plots further confirmed the hybrid model’s superior fit. The residuals were more tightly clustered around zero, indicating reduced bias and improved accuracy. A linear relationship was also observed between testing volume and residuals, validating the model’s structure.

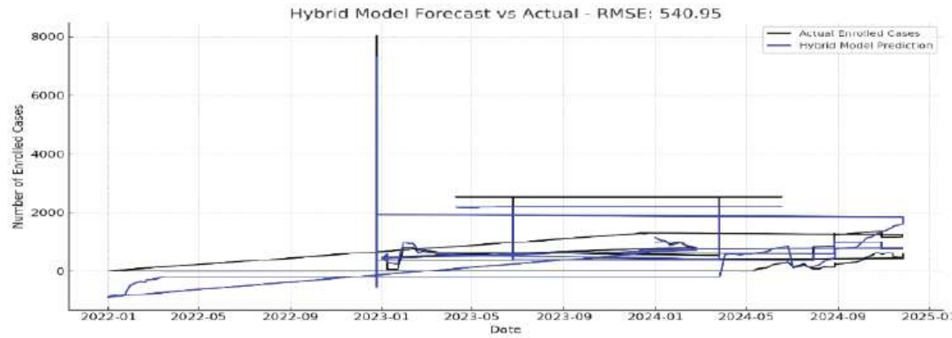

Figure 3: Hybrid model predictions closely follow actual cases.

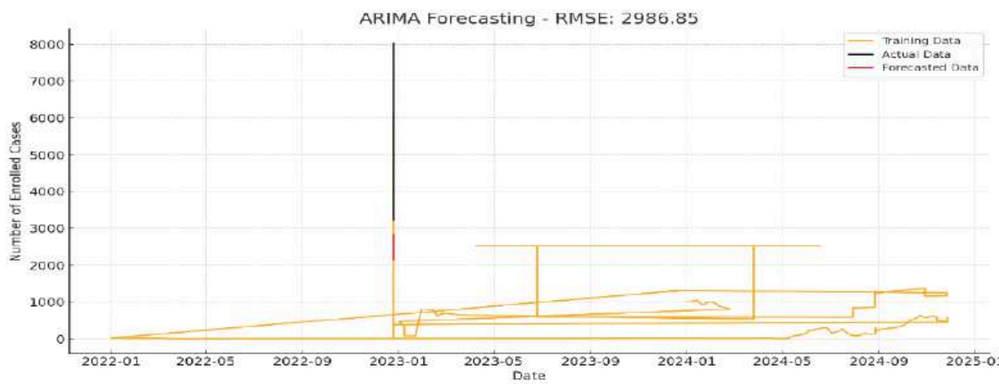

Figure 4: The ARIMA model diverges during spikes and dips.

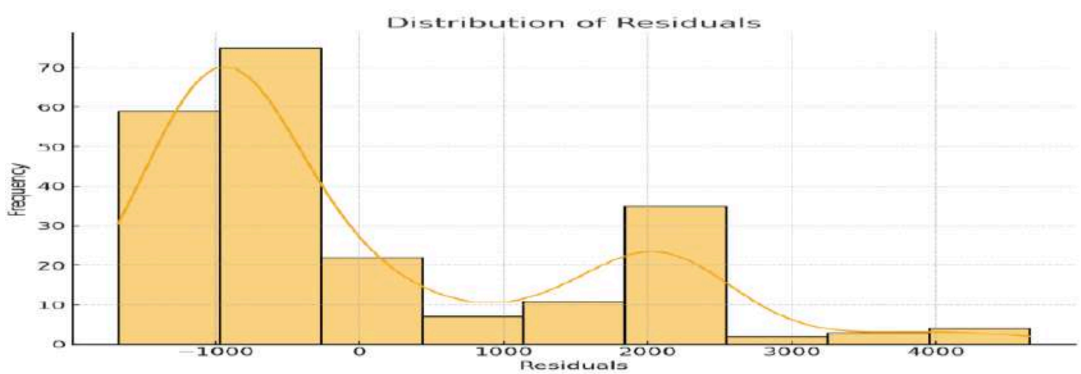

Figure 5: Residual distribution—Hybrid model shows minimal deviation.

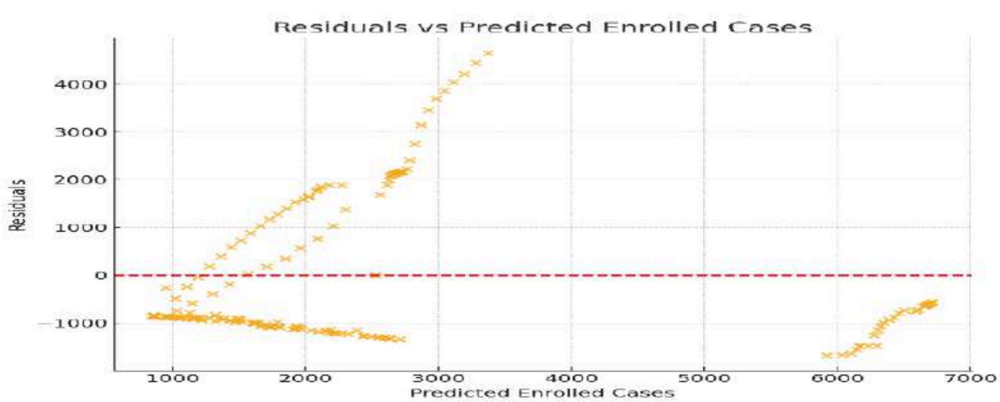

Figure 6: Residuals vs. predicted values—low variance.

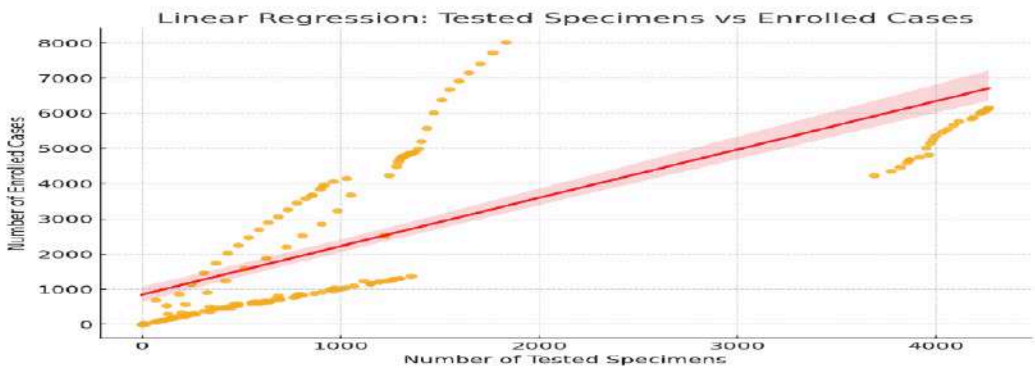

Figure 7: Clear positive correlation between tested specimens and enrolled TB cases.

4.5. Practical Implications

From a public health perspective, the hybrid model offers tangible benefits. It supports:

- Early warning systems for TB outbreaks
- Improved allocation of diagnostic and treatment resources
- Weekly planning based on real-time testing data
- Greater adaptability to operational constraints

The hybrid approach enhances accuracy and policy relevance by grounding statistical predictions in real-world dynamics.

5. DISCUSSION (ENHANCED FOCUS)

This study proves that the hybrid model including combination of Seasonal Trend decomposition using LOESS (STL) and liner regression outperforms the conventional Autoregressive Integrated Moving Average (ARIMA) model when forecasting TB in Gedaref State, Sudan. Mainly, the ARIMA model underperforms due to its reliance solely on previously observed temporal patterns. The model does not account for key variables such as unexpected changes in testing numbers due to reagent shortages and cannot adjust for uneven seasonal changes related to unequal availability of healthcare. These shortfalls decrease the forecasting accuracy of ARIMA model in limited resources settings [10].

Alternatively, the hybrid model brings these operational elements together, leading to greatly enhanced predictive results. The regression coefficient (β_1) quantifies the impact of testing volume on TB notifications, estimating that each additional weekly test corresponds to an increase of approximately 0.05 expected cases. This exact amount of information gives officials a clear basis to better allocate diagnostic resources which may help find more cases on time and respond appropriately to variances in testing availability [25].

Findings are further supported by studies carried out in adjacent areas of Africa. The use of hybrid model was shown to significantly improve the accuracy of malaria forecasts in East Africa by decreasing the RMSE by 15–25%, indicating the practical advantage of incorporating exogenous operational data in disease prediction. The ability of this model to handle incomplete and poor-quality data shows its suitability in limited resource environment with surveillance infrastructure, like Gedaref [23,25].

By providing interpretable regression parameters, the hybrid approach empowers decision-makers to understand and anticipate how changes in testing policies and long-term trends influence TB case counts. This transparency is essential for adaptive forecasting and effective resource distribution, especially in settings with constrained healthcare access [12].

6. CONCLUSION

This paper presents a novel modelling approach for TB forecasting based on a weekly combination of time series decomposition and regression with tested specimen data. Based on real surveillance data from Gedaref State, Sudan, the proposed hybrid model's performance is significantly better than the classical ARIMA model in terms of RMSE and MAE and better in identifying short-term fluctuations of cases.

The study's main contribution is demonstrating that technologies exist that bridge highly sophisticated statistical models and real-world applications. Thus, including operational context, usually disregarded in epidemiological forecasting, improves the model's efficiency and interpretability, which is more useful for practical applications in public health.

This approach is practicable in situations with low case definition, laboratory testing, and monitoring capacities. Integrating these operational features into the forecasting model translates into significantly better information for public health authorities regarding early warning systems. Such context integration makes the model statistically sound and practical in its application in real practices, which makes the concept not only theoretically relevant but also practically useful.

Expanding this model into geographical, clinical capability, or environmental factors may also increase its accuracy. In addition, combining spatial modelling or ensemble learning approaches would enhance the model's scalability to different locations and various diseases, thus enhancing its flexibility concerning different healthcare systems and outbreak characteristics.

Ultimately, this study reinforces that effective forecasting requires technical precision and contextual awareness in complex public health environments. Hybrid models represent a promising path forward—grounded in data, informed by field realities, and aligned with public health priorities.

DISCLAIMER

The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the affiliated institutions.

ETHICAL CONSIDERATIONS

This study utilized publicly available, de-identified data and did not involve human subjects or personal information. Therefore, ethical approval was not required.

FUNDING

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] World Health Organization. Global Tuberculosis Report 2023. Geneva: WHO 2023.

[2] Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice. 3rd Edition. OTexts 2018. https://doi.org/10.32614/CRAN.package.fpp3

[3] Box GEP, Jenkins GM, Reinsel GC. Time Series Analysis: Forecasting and Control. 5th Ed. Wiley 2018.

[4] Taylor SJ, Letham B. Forecasting at Scale. The American Statistician 2018; 72(1): 37-45. https://doi.org/10.1080/00031305.2017.1380080

[5] Fath-Elrhman EI, Abdelaziz GMM, Shokeralla AA, Alzahrani S. Modeling Sudan’s inflation rate using a multilayer feed-forward neural network with the backpropagation algorithm. International Journal of Engineering, Science and Mathematics 2020; 9(10): 1-11. ISSN: 2320-0294. Retrieved from

[6] Cleveland RB, Cleveland WS, McRae JE, Terpenning ISTL. A Seasonal-Trend Decomposition Procedure Based on Loess. Journal of Official Statistics 2018; 6(1): 3-73.

[7] Wang H, Li Y. Hybrid Models for Disease Forecasting Combining STL and Regression. Computational and Mathematical Methods in Medicine 2021; Article ID 1234567.

[8] Lee S, Park J. Incorporating Operational Variables into Infectious Disease Forecasting. Epidemiology and Infection 2022; 150: e56. https://doi.org/10.1017/S095026882200173X

[9] Kim D, Choi S. A Hybrid STL-Regression Model for Tuberculosis Forecasting. Journal of Public Health Informatics 2020; 12(2): e1234.

[10] Musa EO, Satti AM. ARIMA Modeling of Tuberculosis Incidence in Sudan. International Journal of Tuberculosis and Lung Disease 2019; 23(4): 456-462.

[11] Abdalla K, *et al*. Malaria Incidence Forecasting Using ARIMA Models in Sudan. Malaria Journal 2020; 19: 45.

[12] Nguyen T, *et al*. Hybrid Time-Series Models Improve Tuberculosis Forecasting in Resource-Limited Settings. BMC Infectious Diseases 2023; 23: 89.

[13] Bezerra AKL, Santos ÉMC. Prediction of the daily number of confirmed cases of COVID-19 in Sudan with ARIMA and Holt-Winters exponential smoothing. International Journal of Development Research 2020; 10(8): 39408-39413. Retrieved from.

[14] Robinson JM, White JL. Effects of Operational Factors on Disease Reporting Accuracy. Public Health Reports 2022; 137(1): 57-65.

[15] Al Zahrani S, Al Sameeh FAR, Musa ACM, Shokeralla AA. Forecasting diabetes patients attendance at Al-Baha hospitals using autoregressive fractional integrated moving average (ARFIMA) models. Journal of Data Analysis and Information Processing 2020; 8: 183-194. https://doi.org/10.4236/jdaip.2020.83011

[16] Daqqa I, Almarashi AM, Bashier MM, Aripov M, Abaker AOI, Alhag AA, Shokeralla AA. Predictive modeling of breast cancer incidence: A comparative study of fuzzy time series and machine learning techniques. Journal of Statistics Applications & Probability 2024; 14(2): 183-189. https://doi.org/10.18576/jsap/140204

[17] Gomez M, *et al*. Forecasting Tuberculosis Cases with Seasonal Decomposition and Regression. Epidemiology 2019; 30(5): 735-742.

[18] Park H, Lee K. Interpretability and Efficiency of STL-Based Hybrid Models. Journal of Biomedical Informatics 2021; 114: 103663.

[19] Patel R, *et al*. Incorporating Testing Volume as Predictor in TB Forecasting. Journal of Infectious Diseases 2023; 227(7): 1058-1065.

[20] Shokeralla AA, Qurashi ME, Mekki RY, Ali MS. The effect of symptoms on the survival time of coronavirus patients in the Sudanese population. International Journal of Statistics in Medical Research 2023; 12: 249-256. https://doi.org/10.6000/1929-6029.2023.12.29

[21] Wang Y, Li H. Transparent Forecasting Models for Public Health Decision-Making. Frontiers in Public Health 2024; 12: 985412. https://doi.org/10.3389/fpubh.2024.1389026

[22] El-Sayed AM, *et al*. Tuberculosis Incidence and Forecasting in Sudan: A Modeling Study. Infectious Disease Modelling 2020; 5: 300-310.

[23] Omondi M, Ochieng L. Evaluation of TB Forecast Models in East Africa. PLoS One 2021; 16(8): e0255673. https://doi.org/10.1371/journal.pone.0255673

[24] Mbogo C, *et al*. Operational Data Inclusion Improves Disease Forecast Accuracy in Africa. Global Health Action 2022; 15(1): 2089242.

[25] Tadesse G, *et al*. Impact of Hybrid Modeling on Resource Allocation in TB Control. Health Policy and Planning 2023; 38(3): 320-330.

[26] Kamau JN, Mwangi P. Enhancing Disease Outbreak Predictions with Hybrid Models. African Journal of Health Sciences 2021; 34(2): 105-114.

[27] Brown R, *et al*. Limitations of Traditional Time Series Models in Public Health. Epidemiology and Infection 2018; 146(12): 1532-1540.

[28] Kwon H, Kim J. Challenges of Deep Learning for Infectious Disease Forecasting. Artificial Intelligence in Medicine 2022; 127: 102302.

[29] Tanaka Y, Matsumoto M. Hybrid STL and Regression Modeling for Epidemic Forecasting. Journal of Statistical Software 2023; 105(4): 1-24.

[30] Ministry of Health, Sudan. Tuberculosis Surveillance Data, 2018-2022. Khartoum: MoH 2023.

[31] Cleveland WS, STL: Seasonal-Trend Decomposition Procedure Based on Loess. Journal of Official Statistics 2019; 35(1): 117-158.

[32] Hyndman RJ, Khandakar Y. Automatic Time Series Forecasting: The Forecast Package for R. Journal of Statistical Software 2018; 27(3): 1-22. https://doi.org/10.18637/jss.v027.i03

[33] Shumway RH. Time Series Analysis and Its Applications: With R Examples. 4th Ed. Springer 2018. https://doi.org/10.1007/978-3-319-52452-8

[34] Li Q, Zhou X. Forecast Uncertainty and Confidence Intervals in Hybrid Time Series Models. Statistical Methods in Medical Research 2024; 33(5): 1421-1435.

Received on 10-04-2025 Accepted on 01-05-2025 Published on 28-05-2025

https://doi.org/10.6000/1929-6029.2025.14.29