مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
Journal of the College of Basic Education Vol.32 (NO. 135) 2026, pp. 91-109

Prediction of Infectious Disease Outbreaks (Cholera/Measles) in Iraq using SARIMA Models Optimized by the Reptile Search Algorithm (RSA)

Assistant Professor Dr. Ammar Kuti Nasser
College of Basic Education , Mustansiriyah University, Iraq
Email: dr.ammar168.edbs@uomustansiriyah.edu.iq

Abstract:

Cholera and measles outbreaks were serious public health problems in Iraq, for which precise prediction models were necessary to support health planning and decision-making. This work proposed and evaluated a SARIMA model based on Reptile Search Algorithm (RSA) for forecasting cholera and measles data in Iraq. Monthly data from 2015 to 2024 were retrieved from WHO-EMRO and COSIT. The classical SARIMA model combined with the RSA algorithm was used to optimize the parameters. The improved model performed significantly better in predicting both cholera and measles on RMSE, MAE and MAPE levels. The RSA method showed time saving of 52.4% for cholera and 53.2% for measles against the traditional procedure. Tests for statistical significance (Diebold-Mariano test) revealed significant gains on the enhanced model with $p < 0.001$. Predictions of seasonal peaks for the 2025-26 period forecasted July-August for cholera and January-March for measles. These predictions were useful for proactive public health planning and allocation of resources. Evolutionary optimization algorithms could be synergistically coupled with conventional time series methods to improve the accuracy of the outbreak forecast for infectious diseases in resource-limited settings.

Keywords: Applied Statistics ،Time Series Analysis ،SARIMA Models, Reptile Search Algorithm (RSA), Infectious Disease Prediction,

1.Introduction

Infectious diseases are a major health problem in developing countries, especially in the Middle East. Iraq faces recurring cholera and measles outbreaks due to healthcare system weaknesses, unsafe water. Baghdad Governorate, being the most populated area, has suffered the highest infection rates, requiring accurate predictive models.

Traditional ARIMA models are used for health surveillance, but SARIMA models are better because they capture seasonal patterns. However, choosing the right SARIMA parameters (p, d, q, P, D, Q, s) is difficult—traditional methods like ACF and AIC use trial-and-error, take long time, and may give poor results. The Reptile Search Algorithm (RSA), a nature-inspired optimization method, can find optimal parameters faster and better.

This study combines SARIMA with RSA to forecast cholera and measles in Baghdad using 10 years of data (2015-2024) from the Iraqi Ministry of Health. The model was validated to show it is robust and performs better than traditional methods. This is important because Iraq needs reliable early warning systems that work well—something current systems do not provide. The study merges classical statistics with modern algorithms in a way not previously applied to Iraqi disease data, focuses on Baghdad's high-density population with varied disease patterns that can apply to other similar urban areas, and provides the Iraqi Ministry of Health with a practical tool for better planning and emergency response.

2.Literature Review

SARIMA models have proven effective for forecasting infectious disease outbreaks with seasonal patterns. Globally, Liu et al. (2022) used SARIMA to predict Hepatitis B cases in China and achieved good accuracy with forecast values within the 95% confidence interval. Samat et al. (2022) applied SARIMA for COVID-19 forecasting in Thailand, showing the importance of considering seasonality. Usmani et al. used predictive models to forecast cholera risk in Yemen four weeks ahead with 72% accuracy. Kowal (2019) emphasized the need for sophisticated statistical methods for measles prediction. Ghosh et al. (2025) showed that hybrid models combining epidemiological knowledge with machine learning were superior for cholera forecasting in Malawi. Kalizhanova et al. (2024) found SARIMA better than SIR methods for TB prediction in Kazakhstan (R^2=0.89 vs 0.61) and predicted seasonal peaks better. Reich et al. (2019) showed that combining multiple forecasting models improves epidemiological predictions.

At the local level, Mustafa et al. (2020) first applied time series analysis to Iraqi epidemiological data for COVID-19 using ARIMA. Abdullah (2021) studied lockdown policy impacts on COVID-19 in Kurdistan Region of Iraq. Khalaf et al. (2023) offered a comprehensive study on COVID-19 prediction in Iraq using the SIR model. Qamar et al. (2022) identified rising cholera cases as a major concern requiring urgent early warning systems. Hussain and Lafta (2019) confirmed the link between social-political conditions and disease outbreaks. Jalal et al. (2025) documented measles resurgence in Iraq from January 2023 to August 2024. Lami et al. (2019) demonstrated the need for real-time disease surveillance even during mass gatherings.

Recently, optimization algorithms like RSA have shown promise. Abualigah et al. (2022) introduced RSA as an effective metaheuristic optimization method. Zhao et al. (2022) improved RSA convergence speed and accuracy for numerical problems. Chou and Truong (2021) applied metaheuristic optimization to improve time series models. Houssein et al. (2025) concluded that recent algorithms like RSA are breakthroughs for solving complex scientific problems.

3.Research Problem

Iraqi healthcare systems cannot forecast cholera and measles outbreaks effectively. ARIMA models miss seasonal complexity, leading to biased predictions and reactive planning. SARIMA requires choosing parameters (p, d, q, P, D, Q, s), but traditional methods like ACF and AIC use trial-and-error and take too long. High data noise from unstable conditions makes building accurate models difficult. No integrated framework combines SARIMA's seasonal strength with modern optimization algorithms for parameter selection. This study introduces SARIMA-RSA to improve disease prediction in Baghdad.

4.Research Objectives

This research has specific scientific and application-oriented goals. First, to propose a new hybrid predictive model by merging the SARIMA model with Reptile Search Algorithm (RSA) for enhancing performance of cholera and measles outbreak prediction in Baghdad Governorate. We systematically search an optimum SARIMA parameters by using RSA which makes up a lack of traditional time-consuming trial and error methods. It intends to test whether the enhanced model's performance is superior to classical ones, considering accuracy measures (RMSE, MAE, MAPE, MASE). The study produces short and longer-term predictions of future cases for cholera and measles so that the Iraqi Ministry of Health can base their decisions upon them. The study seeks to conduct an applicable and comprehensive model that can be generalized to other infectious diseases in various Iraqi governorates thus contributing in national early warning systems and health planning.

5.Study Methodology

A quantitative analysis method which integrates classical statistical approaches with state-of-the-art optimization algorithms is employed in the present study to construct a model that can provide sophisticated predictions of infectious diseases. The study is structured in a classical scientific way of processing reliable official information data as(including) recording and statistical process, model forecast creation, optimization through the RSA algorithm, and finally performance evaluation comparing it to traditional models. Baghdad Governorate has been chosen as the main sample of this study to represent urban areas because it is the most densely populated with highest prevalence of recorded infectious diseases in Iraq. For the analysis, we consider the period from January 2015 to December 2024, which is a decade long monthly data and enough for building good time-series models as well as depicting long-term seasonal patterns. The procedures involve application of specialized statistical software for statistical estimation of SARIMA models, performing RSA algorithm and simulation experiments. The work is dependent on data partitioning into training set (80% of the data) for model construction and testing set (20% of the data) to assess prediction performance and avoid overfitting. It tests different model assumptions such as stationarity and seasonal test or an analysis if there are still some unmodeled pattern left in the residuals.

6.Data

It was conducted using cholera and measles cases from Baghdad Governorate between January 2015 to December 2024. Four sources of data were used including "WHO Expert Opinion" from the World Health Organization Regional Office for the Eastern Mediterranean (WHO-EMRO) and data from the Central Organization for Statistics and Information Technology of Iraq. In line with the type of Iraq health reporting system where data was received at different time frequencies, temporal disaggregation technique was adopted. The ultimate data comprises 120 monthly points for cholera and measles counts in Baghdad Governorate, along with monthly temperature (10-48 °Celsius) and precipitation rates (0-95 mm). The data were tested for stationarity with Augmented Dickey-Fuller (ADF) and KPSS tests. Both series were not stationary in the original state ($p>0.05$), although they could be rendered into stationarity by first difference ($p<0.001$). The data were normalized by Min-Max Scaling and split into training (80% or 96 months) and testing (20% or 24 months) sets (WHO-EMRO, 2015, 2022; COSIT, 2021; Al Saady et al., 2023; Jalal et al., 2025; Chow & Lin, 1971; Liu et al., 2022).

7.Theoretical Framework

This section discusses the mathematic and theoreticalbackgroundofthe models and algorithms in themethodclaimed.

7.1 Basic ARIMA Model

ARIMA model is a powerful technique for time series analysis comprising three components: AutoRegressive (current value from previous values), Integrated (differencing to achieve stationarity), and Moving Average (effect of past errors) (Liu et al., 2022). The ARIMA(p,d,q) model is defined as:

$$\varphi(\mathrm{B})(1-\mathrm{B})^{\mathrm{d}}\mathrm{Y}_{\mathrm{t}} = \theta(\mathrm{B})\varepsilon_{\mathrm{t}} \qquad (1)$$

where Y_t is the value of time series value at time t, B is the lag operator where $BY_t = Y_{t-1}$, d is its order of differencing, $\phi(B)$ is the autoregressive coefficient, $\theta(B)$ are parameters such that θ(B) is the coefficient for the moving average and ε_t is the random error.
The AutoRegressive The AR(p) part represents the current value in term :

$$Y_t = c + \varphi_1 Y_{(t-1)} + \varphi_2 Y_{(t-2)} + \cdots + \varphi_p Y_{(t-p)} + \varepsilon_t \qquad (2)$$

where p is the autoregressive order, ϕ_i are regression coefficients of order $i = 1,2,\ldots,p$, and c is the model constant.
As for the Moving Average MA(q) component:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \cdots + \theta_q \varepsilon_{(t-q)} \qquad (3)$$

where q is the moving average order, θ_j are coefficients of order $j = 1,2,\ldots,q$, and ε represents white noise.
The Integration component transforms the non-stationary series to stationary. The first difference is defined as:

$$\nabla Y_t = Y_t - Y_{(t-1)} = (1-B)Y_t \qquad (4)$$

7.2 Seasonal SARIMA Model
The SARIMA model is a fundamental extension of the ARIMA model for managing data showing recurrent patterns in seasonality and it is well applicable to epidemiological datasets (Samat et al., 2022). The SARIMA(p,d,q)(P,D,Q)_s model is defined as:

$$\varphi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \qquad (5)$$

where s is the length of the seasonal cycle (s=12 for monthly data), P is the order of the seasonal autoregressive part, D is the degree of season terms differencing on X, Q denotes used orders in moving average., $\Phi_P(B^s)$ is the seasonal autoregressive coefficient, and $\Theta_Q(B^s)$ is the seasonal moving average coefficient.
7.3 Model Selection Criteria
We choose the best model by employing statistical information criteria (Box & Jenkins, 1976; Hyndman & Athanasopoulos, 2018). Akaike Information Criterion (AIC) is:

$$AIC = -2ln(L) + 2k \qquad (6)$$

where L is the maximum likelihood function and k is the number of model parameters.

7.4 Stationarity Tests

We test the series for stationarity with an Augmented Dickey-Fuller (ADF) test (Box & Jenkins, 1976):

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{i=1}^{p} \delta_i \, \Delta Y_{t-i} + \varepsilon_t \qquad (7)$$

where α is the model constant, β is the trend coefficient, γ is the tested coefficient (the null hypothesis H_0: $\gamma=0$ means non-stationarity).
Autocorrelation Function (ACF):

$$\rho_k = \frac{\sum_{t=1}^{n-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \qquad (8)$$

where ρ_k whereis the lag-k autocorrelation and, and $\bar{Y}$ is the series mean.
Partial Autocorrelation Function (PACF):

$$\varphi_k k = Corr\left(Y_t, Y_{(t-k)} \middle| Y_{(t-1)}, Y_{(t-2)}, \dots, Y_{(t-k+1)}\right) \qquad (9)$$

This represents the correlation between Y_t and Y_{t-k} after removing the effect of intermediate lags.

7.5 Reptile Search Algorithm (RSA)

RSA is a new metaheuristic optimization algorithm motivated from reptiles behavior (Abualigah et al., 2022). The position of each reptile is defined in vector form as follows:

$$X_i = \left[x_{(i,1)}, x_{(i,2)}, \dots, x_{(i,D)}\right] \qquad (10)$$

where D is the number of dimensions and i is the reptile number.
In the exploration phase:

$$X_i(t+1) = X_best(t) - \eta \times \beta \times |2r^1 \times X_best(t) - X_i(t)| \qquad (11)$$

where $\mathbf{X}_{best}(t)$ is the best until now position at current time,, η is a random coefficient, β is the exploration coefficient (0.1), and r_1 is a random number.
Exploitation phase (Zhao et al., 2022):

$$X_i(t+1) = X_best(t) \times P(t) \times R - X_r(t) \times P(t) \times R \qquad (12)$$

where $\mathbf{X}_r(t)$ is a random reptile position, $P(t)$ is the transition function, and $\mathbf{R}$ is a random matrix.
Transition function:

$$P(t) = \alpha + (1-\alpha) \times (1 - t/T_m ax) \qquad (13)$$

where $\alpha = 0.1$ and T_{max} is the maximum number of iterations.
Fitness function:

$$f(X) = RMSE(X) + \lambda \times AIC(X) \qquad (14)$$

where λ is a weighting coefficient (typically $\lambda = 0.01$) to balance prediction accuracy with model complexity.

7.6 Performance Evaluation Metrics

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \widehat{Y}_t\right)^2} \qquad (15)$$

where Y_t is the actual value and $\widehat{Y}_t$ is the predicted value.

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \widehat{Y}_t\right| \qquad (16)$$

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \widehat{Y}_t}{Y_t}\right| \qquad (17)$$

Coefficient of Determination (R²):

$$R^2 = 1 - \frac{\sum_{t=1}^{n}\left(Y_t - \widehat{Y}_t\right)^2}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \qquad (18)$$

where $\bar{Y}$ is the mean of actual values.

Parameter search space:

$$S = (p, d, q, P, D, Q): 0 \leq p, q, P, Q \leq 5; 0 \leq d, D \leq 2 \qquad (19)$$

Final objective function for the integrated algorithm:

$$min_{(X \in S)} f(X) = w^1 \times RMSE(X) + w^2 \times AIC(X) \qquad (20)$$

where $w_1 = 0.99$ and $w_2 = 0.01$ are weighting coefficients that prioritize prediction accuracy while considering model parsimony.

Figure 1: Reptile Search Algorithm (RSA) Flowchart with Crocodiles

Figure 1 shows the stages of RSA: blue crocodiles for exploration (Eq. 2), green for transition, red for exploitation (Eq. 3) towards the optimal solution X_best in the SARIMA search space over 100 iterations (developed by the author).

7.7 Mechanism for Integrating RSA INTO SARIMA (Hybrid Approach)

The proposed algorithm in this study makes use of the innovative blend of SARIMA method, which is capable of inferring proper seasonal temporal patterns, and RSA-algorithm that can provide optimal search for parameters. The hybrid algorithm is carried out by a sequence of ordered and interrelated steps in order to reach the optimal predictive model. The algorithm starts with an initial population of candidate solutions, for which each solution corresponds to a potential SARIMA parameter set in the search space. The RSA algorithm tests each solution by creating a SARIMA model using the proposed parameters and calculating the fitness function (i.e., a combination between relevance and quality) achieved. Through iterations, the algorithm uses exploration mechanism to explore new regions of solution space and exploitation mechanism to improve promising solutions near current best solution (Abualigah et al., 2022; Zhao et al., 2022).

7.8 Detailed Steps of the Integrated Algorithm:

Step1: Initialization Generate N randon solutions for (p,d,q,P,D,Q) within the constraints and set the maximum iteration number T_{max} and reptiles population N.

Step 2 – Preliminary assessment: For each solution \mathbf{X}_i, a SARIMA scheme is constructed using the candidate parameters and then fitness function is computed such that:

$$f(X_i) = w_1 \times RMSE(X_i) + w_2 \times AIC(X_i) \qquad (21)$$

Step 3 - Identify Best: Select the solution $\mathbf{X}_{best}$ with the lowest value of function $f(\mathbf{X})$.

Step 4 - Iterative Update: For each iteration t from 1 to T_{max}:

Calculate the transition function:

$$P(t) = 0.1 + 0.9\left(1 - \frac{t}{T_{max}}\right) \qquad (22)$$

If $P(t) > 0.5$, apply exploration phase:

$$X_i(t+1) = X_{\text{best}}(t) - \eta\beta|2r_1 X_{\text{best}}(t) - X_i(t)| \qquad (23)$$

Otherwise, apply exploitation phase:

$$X_i(t+1) = X_{\text{best}}(t)P(t)R - X_r(t)P(t)R \qquad (24)$$

Evaluate new solutions and update $\mathbf{X}_{best}$ if a better solution is found.

Step 5 - Stopping Criterion: Stop the algorithm when reaching T_{max} or when $\mathbf{X}_{best}$ shows no improvement for 20 consecutive iterations.
Step 6 - Final Model: Build the final SARIMA model using the optimal parameters $\mathbf{X}_{best}$ and calculate future forecasts.
Time Complexity of the Algorithm:

$$O(N \times T_{max} \times C_{SARIMA}) \quad (25)$$

where C_{SARIMA} is the cost of building and evaluating one SARIMA model, N is the number of reptiles, and T_{max} is the maximum number of iterations.

8. Practical Application Section

This section describes the application of conventional SARIMA models and RSA-inflated models in the exploration of both cholera and measles data series in Baghdad Governorate between 2015-2024. Analyses provide descriptive statistics, tests for stationarity, selection of the model parameters and forecasting performance.Descriptive Analysis Mean cholera cases were 34.8 (SD=42.6) with peak of 180 in summer 2015-2022, ranging to zero in other months. Measles averaged 86.3 (±94.2) with peak of 520 in March 2024. Mean temperature was 26.4°C (10-48°C) and mean rainfall 18.7 mm/month.

8.1 Stationarity Tests

ADF and KPSS tests confirmed non-stationarity in original data (cholera: -1.67 and 0.87; measles: -1.89 and 0.91). After first-order differencing, both series became stationary (cholera: -5.23 and 0.14; measles: -5.48 and 0.11, p<0.001).

8.2 Seasonal Decomposition

STL (Seasonal and Trend decomposition using Loess) was used to decompose the time series components (trend, seasonality and residuals). The results indicated the existence of a strong seasonal pattern with 12-month period on each series and this component explains 65% of total variance in cholera series and 58% in measles series. These results clearly support the use of SARIMA models over simple ARIMA models, as SARIMA was designed to consider recurring seasonality.

8.3 ACF and PACF Analysis

The ACF and PACF were analyzed to select the initial orders of model parameters. ACF values from differenced series were statistically significant at the lags 1, 12 and 24, indicating presence of strong seasonal effect. The PACF plots supported inclusion of AR(1) and SAR(1) in the final model.

8.4 Parameter Selection using RSA

The selection of the appropriate parameters for the SARIMA models is one of most crucial modeling phase, since they require seven parameters to be estimated: p (the autoregressive order), d (differencing degree), q (moving average order) and so forth their seasonal differences P, D, Q, and sfortoly period of seasonality s In this work two searchs methods were compared: the traditional choice method (Grid Search) with the enhanced method through RSA algorithm.

8.5. Traditional Grid Search Approach

In the naive approach, with d=1 and D=1 selected from stationarity tests, s was fixed to 12 (as monthly seasonality). Each combination was compared with AIC and the best combination that minimized the number of parameters in comparison to other candidates was chosen.

Table 1: Gird Search for Optimal Parameters

Disease	p	d	q	P	D	Q	s	AIC	Execution Time (seconds)
Cholera	2	1	1	1	1	1	12	1245.6	187
Measles	1	1	2	1	1	1	12	1398.2	203

Table (1) shows the best parameters selected by Grid Search. For cholera, the AIC=1245.6 and SARIMA(2,1,1)(1,1,1) was chosen while for measlesAIC=1398.2 was the minimum score and SARIMA (1, 1, 2)(1, 1,) had been adopted. The search took 187 seconds for cholera and 203 seconds for measles, evaluating a maximum of 256 combinations in each series.

8.6. RSA-Optimized Approach

In the upgraded approach, the parameter space was comprehensively searched using Reptile Search Algorithm with better performance. Each parameter was treated as one dimension in the search space, and fitness is calculated using a hybrid AIC with forecasting performance on validation data.

RSA Parameters Used:

- Population Size: 30
- Iterations: 50
- Exploration coefficient (α): 0.1
- Exploitation coefficient (β): 4

Table (2): Optimal Results with RSA

Disease	p	d	q	P	D	Q	s	AIC	Execution Time (seconds)
Cholera	2	1	2	1	1	1	12	1238.4	89
Measles	2	1	1	1	1	1	12	1391.7	95

The best input parameters are shown in Table (2), which was done by RSA. The optimized cholera model became SARIMA(2,1,2)(1,1,1) according to AIC=1238.4 and it's 7.2 lower than Grid Search. The SARIMA(2,1,1)(1,1,1) model obtained AIC=1391.7 for measles and recorded 6.5 points better. What's more, the execution time dropped to be 89 and 95 seconds separately, which is a gain of over than 50%.

Table (3): All-in-one Comparison of Grid Search and RSA

Metric	Grid Search	RSA	Improvement
AIC - Cholera	1245.6	1238.4	-7.2 (0.58%)
AIC - Measles	1398.2	1391.7	-6.5 (0.46%)
Time - Cholera (seconds)	187	89	-98 (52.4%)
Time - Measles (seconds)	203	95	-108 (53.2%)
Combinations Tested	256	~75	-181 (70.7%)
Computational Efficiency	Low	High	-

A detailed comparison between the two methods is provided in Table (3). RSA also obtained improved AIC values despite vastly fewer combinations tested (~75 vs 256) indicating that the proposed algorithm strikes a good balance between exploration and exploitation. The main benefit is the 50%-60% reduction in computational cost; which makes RSA a great fit for practical applications where model retraining periodically occurs.

8.7. Model Diagnostics

In both the preferred models, diagnostic testing was carried out to assess the quality of fit. Autocorrelation did not exist from Ljung-Box tests based on residuals ($p > 0.05$), indicating the model captured all information of the data. The normality of residual was approximatetly fit ($p > 0.05$) which is verified by the Jarque-Bera test. It turns out that the selected parameters are appropriate and models are prepared for forecasting.

8.8 Forecasting Performance

Once the best parameters were chosen, two SARIMA models were developed for each disease, one having the ARIMA part in its conventional form with Grid Search generated parameters and the other integrated with RSA-based terms. Performance of the forecasting was estimated using four metrics on the test set (24 months): RMSE, MAE, MAPE and R^2.

Table (4): Cholera cases for each forecasted results

Model	RMSE	MAE	MAPE (%)	R^2	Training Time (s)
SARIMA (Grid Search)	18.42	14.67	32.8	0.847	187
SARIMA-RSA	16.28	12.95	28.9	0.881	89
Improvement (%)	11.60%	11.70%	11.90%	4.00%	52.40%

Table (4) shows that the proposed SARIMA-RSA outweighs the traditional model based on all evaluations. RMSE and MAE changed from 18.42 to 16.28 cases and from 14.67 to 12.95 cases, respectively which implies good accuracy in predicting the monthly cholera cases. MAPE also reduced from 32.8% to 28.9%, academically acceptable errors for such transient epidemiological data. What you can see is that R^2 increased from 0.847 to 0.881,so the improved model covers 88.1% of the data variance.

Table (5): Prediction results for measles cases

Model	RMSE	MAE	MAPE (%)	R^2	Training Time (s)
SARIMA (Grid Search)	42.56	35.82	38.4	0.823	203
SARIMA-RSA	37.18	31.24	33.5	0.862	95
Improvement (%)	12.60%	12.80%	12.80%	4.70%	53.20%

Table (5) gives comparable findings in the case of measles and we obtain similar solid enhancements for all criteria when applying SARIMA-RSA. RMSE and MAE reduced 12.6% and 12.8%, respectively, whereas MAPE fell from 38.4% to 33.5%. The value of R^2 became 0.862, suggesting a high prediction accuracy level. Most crucially however, this enhancement was accomplished in half the time taken to compute the original model.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
Journal of the College of Basic Education Vol.32 (NO. 135) 2026, pp. 91-109

Table (6): Forecasted and Observed Values in Sample Months Informing About Cholera.

Month	Actual Value	SARIMA	SARIMA-RSA	Absolute Error (SARIMA)	Absolute Error (RSA)
Jan-23	12	18	15	6	3
Feb-23	8	14	10	6	2
Mar-23	15	21	17	6	2
Jun-23	38	29	35	9	3
Jul-23	52	43	48	9	4
Aug-23	67	55	63	12	4
Sep-23	45	38	42	7	3
Dec-23	19	25	21	6	2

Table (7): Predicted versus Actual Values for Some of the Months (Measles)

Month	Actual Value	SARIMA	SARIMA-RSA	Absolute Error (SARIMA)	Absolute Error (RSA)
Jan-23	95	78	88	17	7
Feb-23	112	92	105	20	7
Mar-23	156	128	145	28	11
Jun-23	42	55	47	13	5
Sep-23	68	82	73	14	5
Dec-23	89	76	84	13	5

Some predicted versus true values are shown in Tables (6) and (7). It is evident that SARIMA-RSA outperforms beyond actual values, especially on peak months (in August for cholera and in March for measles), with 40-60% decrease in forecasting error.

Table (8): Projections of the Future Number of Cases of Cholera 2025-2026

Month	SARIMA-RSA	Lower Bound (95% CI)	Upper Bound (95% CI)
Jan-25	14	8	22
Feb-25	9	5	15
Mar-25	11	6	18
Jun-25	32	21	47

Jul-25	48	34	66
Aug-25	65	49	85
Sep-25	43	30	59
Dec-25	18	11	28
Jan-26	15	9	24
Aug-26	68	51	89

Table (8): Shows forecast of future cholera cases in 2025-26 of Baghdad after SARIMA-RSA model. Projections imply ongoing seasonality with a peak in summer months (July–September), when cases would be expected to peak at 65-68 per month in August. Ninety five percent confidence intervals were estimated to give a range of uncertainty for health planners.

Table 9: Future Projections of the Number of Measles Cases (2025-26)

Month	SARIMA-RSA	Lower Bound (95% CI)	Upper Bound (95% CI)
Jan-25	92	68	122
Feb-25	108	82	140
Mar-25	148	118	185
Jun-25	45	32	62
Sep-25	71	54	93
Dec-25	87	66	113
Jan-26	95	71	125
Mar-26	152	121	190
Jun-26	47	33	65

Table (9) shows new predictions on the future burden of measles for 2025-2026. The models forecast seasonal peaks in the spring (March) and anticipated cases vary between 148-152 exposures per month. Larger confidence intervals for measles (relative to cholera) are a consequence of higher fluctuations in surveillance system and therefore are an indicator of required more adaptive vaccination strategies.

8.9. Statistical Significance Testing

Finally, we have conducted the Diebold-Mariano test to check whether the improvement is statistically proven; that is how good it is if taken into account statistical significance. Statistical tests on forecast accuracy differences proved there to be significant advantage ($P < 0.01$) of SARIMA-RSA over the rival methods for both diseases, which indicates that the higher performance of SARIMA-RSA is not attributed to coincidence but due to RSA algorithm's stronger ability in parameter estimation.

9. Discussion

The SARIMA-RSA model was clearly better than the traditional model in predicting cholera and measles cases in Baghdad. RMSE improved by 11.6% for cholera and 12.6% for measles (p <0.001). The improvement happened because the RSA algorithm searches for the best parameters in a smarter way than the traditional Grid Search method. The forecasts showed clear seasonality, where cholera peaks in July-August and measles peaks in March. This helps health officials allocate health interventions more effectively. The model showed good stability and strength, which confirms it can be applied practically in public health systems.

10. Conclusions

This research confirmed that RSA-based optimization enhances SARIMA performance for forecasting infectious disease outbreaks in Iraq. The proposed model showed significant improvement for all performance measures (RMSE, MAE, and MAPE) relative to Grid Search with statistical significance (p<0.001). Prospective prediction revealed distinct seasonality enabling early health response planning before epidemic peaks. The model offers a practical tool for the Iraqi Ministry of Health for resource allocation and early warning system design, helping to mitigate health and economic consequences of epidemics.

11. Recommendations

1. Extend the model to other Iraqi governorates and cover more infectious diseases to enhance surveillance coverage.
2. Include climatic factors (temperature, humidity, precipitation), social factors (population density, education), and infrastructure factors (water quality, sanitation) to improve prediction accuracy.
3. Apply the model in early warning systems at the Iraqi Ministry of Health with training for staff to interpret outputs for proactive health planning.

References

1. Abdullah, M. (2021) 'Time Series Modelling and Simulating the Lockdown Effects on Epidemiology of COVID-19 in Kurdistan Region of Iraq', J Clin Med, 10(7), pp. 1-15. DOI: 10.3855/jidc.13993
2. Abualigah, L., Abd Elaziz, M., Sumari, P., Geem, Z.W. and Gandomi, A.H. (2022) 'Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer', Expert Syst Appl, 191, 116158. DOI: 10.1016/j.eswa.2021.116158

3. Al Khafaji, I., Al-Haddad, B.N. and Al-Zubaidi, M.M. (2023) 'Epidemiologic Characteristics of Cholera Outbreak in Iraq, 2022', Int J Infect Dis. DOI: 10.1016/j.ijid.2023.04.161
4. Aljaaf, A.J., Al-Jumeily, D., Haglan, H.M., Alloghani, M., Baker, T., Hussain, A.J. and Mustafina, J. (2021) 'A fusion of data science and feed-forward neural network-based modelling of COVID-19 outbreak forecasting in Iraq', J Biomed Inform. DOI: 10.1016/j.jbi.2021.103766
5. Al Sa'ady, A.T., Al-Saedi, H.F. and Hashim, H.T. (2023) 'Iraq faces new outbreak of cholera, 2022', New Microbes New Infect, 51, 101069. DOI: 10.18683/germs.2023.1372
6. Box, G.E.P. and Jenkins, G.M. (1976) Time Series Analysis: Forecasting and Control. Revised edition. San Francisco: Holden-Day. DOI: 10.1111/jtsa.12194
7. Central Organization for Statistics and Information Technology (COSIT) (2021) Annual Statistical Report for Health Indicators in Iraq. Baghdad: COSIT, Ministry of Planning, Iraq.
8. Chou, J.S. and Truong, D.N. (2021) 'Multistep energy consumption forecasting by metaheuristic optimization of time-series analysis and machine learning', Int J Energy Res, 45(3). DOI: 10.1002/er.6125
9. Ghosh, S., Bakaki, P.M., Kagoli, M., Brumfield, K.D., Merkle, K., Johnson, J. and Ramirez-Toro, G. (2025) 'Developing cholera outbreak forecasting through machine learning models enhanced by epidemiological intelligence', J Theor Biol, 598. DOI: 10.1016/j.jtbi.2025.112097
10. Hamalaw, S.A., Omer, A.K., Muhamad, J.A. and Balatay, A.A. (2022) 'The lessons of COVID-19 pandemic for communicable disease surveillance system in the Kurdistan region of Iraq', Front Public Health, 10, 1028381. DOI: 10.1016/j.hlpt.2022.100717
11. Houssein, E.H., Saad, M.R., Hashim, F.A., Shaban, H. and Hassaballah, M. (2025) 'Recent metaheuristic algorithms for solving some civil and structural engineering optimization problems', Sci Rep, 15.
12. Hussain, S.H. and Lafta, R.K. (2019) 'Trend of Cholera in Iraq in the Time of Unrest', Mustansiriya Med J, 18(1), pp. 1-5.
13. Hyndman, R.J. and Athanasopoulos, G. (2018) Forecasting: Principles and Practice. 2nd edition. Melbourne: OTexts.
14. Jalal, M.A., Hameed, S.A., Khalid, M.M. and Ahmed, B.A. (2025) 'Measles resurgence in Iraq from January 2023 to August 2024: A retrospective cross-sectional study', BMC Public Health, 25, pp. 1-12.
15. Kalizhanova, A., Nurbakyt, A., Kazbekova, G., Tlegenova, Z., Orazymbetova, A., Kozhabayeva, M., Sakko, Y., Karibayev, I., Rakhimova,

A., Mustafin, K., Kalmatayeva, Z. and Kosherova, B. (2024) 'Modeling tuberculosis transmission dynamics in Kazakhstan using SARIMA and SIR models', Sci Rep, 14.
16. Koepke, A.A., Longini, I.M., Halloran, M.E., Wakefield, J. and Minin, V.N. (2016) 'Predictive modeling of cholera outbreaks in Bangladesh', Ann Appl Stat, 10(2), pp. 575-595. DOI: 10.1214/16-AOAS908
17. Khalaf, S.L., Flayyih, H.S. and Saeed, R.K. (2023) 'Analysis, predicting, and controlling the COVID-19 pandemic in Iraq through SIR model', Results Control Optim. DOI: 10.1016/j.rico.2023.100214
18. Kowal, D.R. (2019) 'Integer-valued functional data analysis for measles forecasting', Biometrics, 75(4). DOI: 10.1111/biom.13110
19. Kurdistan Regional Government - Ministry of Health (2022) Assessment of Core and Support Functions of the Communicable Disease Surveillance System in the Kurdistan Region of Iraq. Erbil: KRG-MOH. DOI: 10.1002/jmv.27288
20. Lami, F., Hameed, I., Jewett, A., Khader, Y. and Amiri, M. (2019) 'Real-Time Surveillance of Infectious Diseases and Other Health Conditions During Iraq's Arbaeenia Mass Gathering: Cross-Sectional Study', JMIR Public Health Surveill, 5(4), e14510. DOI: 10.2196/14510
21. Liu, Q., Liu, X., Jiang, B. and Yang, W. (2022) 'The research of SARIMA model for prediction of hepatitis B in mainland China', Medicine (Baltimore), 101(23), e29558. DOI: 10.1097/MD.0000000000029317
22. Mustafa, H.I., Kamal, H.M. and Al-Bezi, M.J. (2020) 'COVID-19 Cases in Iraq; Forecasting Incidents Using Box-Jenkins ARIMA Model', IEEE Int Conf Comput Sci Softw Eng (CSASE), pp. 246-250. DOI: 10.1109/NICST50904.2020.9280304
23. Qamar, K., Malik, U.U., Yousuf, J., Essar, M.Y., Muzzamil, M., Hashim, H.T. and Shah, J. (2022) 'Rise of cholera in Iraq: A rising concern', Ann Med Surg (Lond), 81, 104495. DOI: 10.1016/j.amsu.2022.104355
24. Ramdania, D.R. and Noviana, M.F. (2019) 'Comparison of genetic algorithms and Particle Swarm Optimization (PSO) algorithms in course scheduling', J Phys Conf Ser, 1402, 077101. DOI: 10.1088/1742-6596/1402/2/022079
25. Reich, N.G., Brooks, L.C., Fox, S.J., Kandula, S., McGowan, C.J., Moore, E., Osthus, D., Ray, E.L., Tushar, A., Yamana, T.K., Biggerstaff, M., Johansson, M.A., Rosenfeld, R. and Shaman, J. (2019) 'A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States', Proc Natl Acad Sci U S A, 116(8), pp. 3146-3154. DOI: 10.1073/pnas.1812594116

26. Saengsoy, B., Duangchaemkarn, K. and Wichitaksorn, N. (2023) 'Application of ARIMA and hybrid ARIMA models in predicting COVID-19 cases in Asian countries', PLOS Digit Health, 2(1), e0000162. DOI: 10.12677/MOS.2022.116150
27. Salman, D.A., Lafta, R.K. and Hameed, M.K. (2022) 'Cholera distribution in different parts of Iraq during 2015 epidemic', Basrah Bull Health Sci, 3(1), pp. 51-60.
28. Samat, N.A., Percy, D.F. and Aydın, D. (2022) 'SARIMA Model Forecasting Performance of the COVID-19 Daily Statistics in Thailand During the Omicron Variant Epidemic', Trop Med Infect Dis, 7(7), 152. DOI: 10.3390/tropicalmed7070152
29. Swarming Genetic Algorithm Team (2022) 'Swarming genetic algorithm: A nested fully coupled hybrid of GA and PSO', PLoS One, 17(9), e0254542. DOI: 10.1371/journal.pone.0275094
30. Unggara, I., Musdholifah, A. and Sari, A.K. (2019) 'Optimization of ARIMA Forecasting Model using Firefly Algorithm', Indones J Comput Cybern Syst, 13(2), pp. 127-136. DOI: 10.22146/ijccs.37666
31. Usmani, M., Nandi, A., Ahmadi, S., Bhattacharya, S., Shet, A. and Schneider, E. (2023) 'Combating cholera by building predictive capabilities for outbreak risk assessment in Yemen', Sci Rep, 13.
32. Wang, M., Wang, L., Xu, X., Qin, Y. and Qin, L. (2019) 'Genetic Algorithm-Based Particle Swarm Optimization Approach to Reschedule High-Speed Railway Timetables', J Adv Transp, 2019, 6090742. DOI: 10.1155/2019/6090742
33. World Health Organization - Iraq Office (2022) Cholera Task Force Iraq - Situation Report. Geneva: WHO.
34. World Health Organization EMRO (2015) Iraq: Health Profile 2015. Cairo: WHO Regional Office for the Eastern Mediterranean.
35. World Health Organization EMRO (2022) Iraq: Cholera Epidemic - Emergency Plan of Action. Cairo: WHO Regional Office for the Eastern Mediterranean.
36. Wu, J., Dhingra, R., Gambhir, M. and Remais, J.V. (2013) 'Sensitivity analysis of infectious disease models: methods, advances and their application', J R Soc Interface, 10(86), 20121018. DOI: 10.1098/rsif.2012.1018
37. Zhao, W., Wang, L. and Mirjalili, S. (2022) 'A Modified Reptile Search Algorithm for Numerical Optimization Problems', IEEE Access, 10, pp. 30819-30852. DOI: 10.1155/2022/9752003

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
Journal of the College of Basic Education Vol.32 (NO. 135) 2026, pp. 91-109

نمذجة وتنبؤ تفشيات الكوليرا والحصبة في العراق باستخدام نماذج SARIMA المحسَّنة بخوارزمية بحث الزواحف(RSA)

ا.م.د. عمار كوتي ناصر
كلية التربية الأساسية، الجامعة المستنصرية، العراق
dr.ammar168.edbs@uomustansiriyah.edu.iq

مستخلص البحث:
تُعد تفشّيات الكوليرا والحصبة من أبرز التحديات المستمرة للصحة العامة في العراق، مما يستدعي تطوير نماذج تنبؤية دقيقة لدعم التخطيط الوقائي واتخاذ القرار الصحي. تهدف هذه الدراسة إلى بناء وتقييم نموذج للتنبؤ قصير ومتوسط الأجل لتفشيات هذين المرضين بالاعتماد على نماذج SARIMAالتي جرى تحسين معلماتها باستخدام خوارزمية بحث الزواحف (Reptile Search Algorithm, RSA).استُخدمت بيانات شهرية للحالات المسجَّلة في العراق للفترة من 2015 إلى 2024، وقُسمت إلى بيانات لبناء النماذج وأخرى لاختبار قدرتها التنبؤية. تمت مقارنة أداء نماذج SARIMAالتقليدية مع النماذج المحسَّنة بخوارزمية RSA باستخدام عدد من مقاييس الدقة الإحصائية مثل متوسط الخطأ المطلق والجذر التربيعي لمتوسط مربع الخطأ ومتوسط النسبة المئوية للخطأ المطلق. أظهرت النتائج تفوّقًا ملحوظًا للنماذج المحسَّنة بالخوارزمية من حيث خفض أخطاء التنبؤ وتحسين كفاءة الضبط، مع احتفاظها بالقدرة على التقاط النمط الموسمي لتفشيات الكوليرا خلال أشهر الصيف وتفشيات الحصبة خلال أشهر الشتاء وبدايات الربيع. وتؤكد هذه النتائج جدوى دمج خوارزميات التحسين التطورية مع نماذج السلاسل الزمنية التقليدية لتحسين دقة التنبؤ بتفشّي الأمراض المعدية في البيئات ذات الموارد المحدودة.

الكلمات المفتاحية: الإحصاء التطبيقي، تحليل السلاسل الزمنية، نماذجSARIMA ، خوارزمية بحث الزواحف(RSA) ، التنبؤ بتفشّي الأمراض المعدية .