Applied Medical Informatics
Original Research Vol. 47, No. 3/2025, pp: 64-75

Forecasting Scabies Trends in Ghana using Seasonal Autoregressive Integrated Moving Average and Generalized Linear Model

Michael Asante OFOSU [1,*], Emmanuel Owiredu ODAME [1], Magdalene TORTO [2], Mercy Anna NUAMAH [2], Akoto Yaw Omari SASU [1], Wilhemina Adoma PELS [1], Sandra ADDAI-HENNE [1], Princess Sedinam AKORLIE,[1] and Elvis Anyimadu ASIEDU [1]

[1] Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, MCFM+F94, Aboagye Menyah Complex, College of Science (KNUST), Kumasi, Ghana.
[2] Department of Obstetrics and Gynecology, University of Ghana Medical School, Korle-Bu Teaching Hospital, P.O. Box GP 4236, Accra, Ghana.
E-mails: *mike.ofosu16@gmail.com; emmanuel.odame@knust.edu.gh; magdaleneodame@gmail.com; manuamah@ug.edu.gh; ayos@nims.edu.gh; wilheminaadoma.pels@knust.edu.gh; addaihenne@gmail.com; akorlieprincesssedinam9@gmail.com; elvisasiedu2002@gmail.com

* Author to whom correspondence should be addressed;

Received: 24 January 2025 /Accepted: 9 September 2025/ Published online: 27 September 2025

Abstract
Background: Scabies, a neglected tropical disease, poses a significant public health challenge in resource-limited settings such as Kokofu in Ghana's Ashanti Region. Accurate forecasting of scabies incidence is crucial for effective allocation and management of healthcare resources. *Objective:* This study aimed to develop and validate predictive models of scabies incidence in the Ashanti Region of Ghana using time series analysis and generalized linear models. The goal was to demonstrate the potential utility of these models in enhancing infectious disease control by generating accurate predictions to inform healthcare resource allocation and planning. *Methods:* Monthly scabies case data from January 2016 to May 2023 were extracted from the Kokofu Hospital Information System. We applied Seasonal Autoregressive Integrated Moving Average (SARIMA) models for time-series analysis and compared Poisson and Negative Binomial regression models. The model selection was based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Forecasting accuracy was assessed using Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE). *Results:* The SARIMA(1,1,2)(0,1,1) modelhas emerged as the optimal forecasting tool (AIC: 866.79, BIC: 878.44). This model revealed a significant upward trend in scabies incidence and distinct seasonal patterns, with peak incidence during the dry season. The Negative Binomial regression outperformed the Poisson model (AIC: 1039.6 vs. 2496.4), identifying significant increases in scabies cases in June and July; and decreases in November and December. *Conclusion:* Our study demonstrated the efficacy of combining time-series analysis with generalized linear models in forecasting the incidence of scabies. The developed models provide a robust framework for predicting outbreaks and seasonal variations, offering valuable insights into healthcare planning and resource allocation.Our approach has the potential to enhance scabies management and can be integrated into existing health information systems to support evidence-based decision making in similar resource-limited settings.

Keywords: Scabies; Time Series Analysis; Linear Models; Public Health.

Introduction

Human skin is the largest organ in the body and plays an important role in the body and the external environment. The skin is affected by various chemicals, such as allergens, irritants, and pollutants, as well as living organisms like bacteria, fungi, and parasites [1]. Poisons from foods and medications can also cause skin irritation and infections. Scabies are very common in many parts of the world, particularly in low-income tropical regions, with prevalence rates ranging from 0.2% to 71.4% depending on the population studied [2]. In temperate regions, scabies has a higher impact in winter than in summer, with studies showing a 20-30% increase in cases during colder months [3].

Scabies is caused by a tiny mite called *Sarcoptes scabiei*, which burrows into the upper layer of the skin and lays its eggs [4]. Female mites burrow into the skin and cause scabies, which is caused by mites in human hosts. The female burrows into the skin and lays eggs, which develop into larvae in 2 to 3 days and into mites in approximately 15 days. Normal scabies usually has five to fifteen female mites, whereas crusted scabies can have hundreds or even millions of mites [5]. Scabies causes itching and inflammation of the skin due to mite infestation [6]. Scabies can be transmitted through close physical contact, including sexual activity, as mites that cause scabies spread through prolonged skin-to-skin contact [7]. Although it is rare for scabies to spread through shared clothing or other indirect exposures, it can occur during normal infections. This can cause skin irritation and serious complications such as sepsis (blood infection), heart disease, and kidney problems.

Scabies disease not only affects physical health but also mental health. The physical discomfort and social stigma associated with the condition can lead to embarrassment, social isolation, anxiety about contagion, low self-esteem, and fear of re-infestation. These mental health impacts were experienced by 77% of those affected [7, 3]. Scabies mainly affect low-income tropical areas, increasing the vulnerability of children and the elderly in areas with limited resources [3]. According to a British epidemiological study (United Kingdom, 2014), scabies has a higher impact in winter than in summer, with a 20-30% increase in cases during colder months, especially among women and children [3]. The greatest impact of the disease is seen in low- and middle-income countries, where overpopulation and lack of access to quality healthcare play a major role in promoting transmission [8].

Romani et al. [2] estimated the global scabies prevalence to be between 0.2% and 71.4%, depending on the populations studied. In Ghana, a retrospective study of skin diseases at Korle Bu Teaching Hospital identified a scabies disease rate of 5.1% [9]. However, the incidence of scabies across countries has not yet been systematically reported. Information on scabies, both in general and within published sources, is limited. Despite scabies being acknowledged as a common skin ailment in numerous impoverished rural and urban communities in Ghana, the available evidence primarily comprises unreliable accounts. In general, scabies exert a greater impact in hot tropical areas, where infestations are persistent and intermittent outbreaks occur [10]. Although there are effective treatments, scabies is still classified as a tropical disease because there is no systematic approach to its treatment [11].

The prevalence of scabies is very high in some countries, ranging from 32.1% to 74%, especially in crowded places such as prisons, boarding schools, and orphan care centers [8]. In September 2019, an outbreak of the disease was recorded in locations such as the East Mamprusi region of northern Ghana [8]. Fortunately, the World Health Organization (WHO) formally identified scabies as neglected tropical disease in 2017 to raise awareness and boost efforts to eliminate it [13]. Scabies is also included in the WHO Roadmap for Neglected Tropical Diseases 2021-2030, which aims to end neglect to achieve Sustainable Development Goals. Among these, scabies is a major disease [14]. To prevent and control the disease, it is important to have knowledge of the disease and develop an effective prevention and control plan.

Recent studies have also applied SARIMA models to scabies incidence, such as Ghiebi et al. [15], who analyzed vector-borne diseases including scabies in Iran, and Perez et al. [16], who studied scabies infections among wild ibex in Spain. These studies highlight the utility of time-series models in understanding scabies dynamics across different contexts.

The present study aimed to apply a time series and generalized linear model to forecast and identify months with infections contributing to scabies diseases in the Bekwai Local Government Area of the Ashanti Region, Ghana.

Materials and Methods

Data Source and Study Design

The data used in this research was secondary and consisted of monthly scabies case records extracted from the Kokofu Hospital Information System, covering the period from January 2016 to May 2023. The data were obtained with permission from the hospital administration and is not publicly available due to patient confidentiality agreements

Time Series Model

Time series analysis is a statistical method used to analyze and interpret data points collected over a specific time, typically at regular intervals [17]. The components of the time series refer to the distinct underlying patterns of variation within a time series, that is, trend, seasonal, cyclic, and irregular components [18]. A time series is generally considered stationary if periodic variation is eliminated and there is no change in the mean (no trend) or variance. Metrics like the mean and variance change over time and revert to long-term averages [19]. Several models are commonly used in time-series analyses, and the features of the data mostly influence model selection. The autoregressive (AR) model of order p, denoted as AR(p), expresses the current value of a time series Y_t as a linear function of its past p-values. It is defined as:

$$Y_t = \sum_{h=1}^{p} \phi_h Y_{t-h} + \mu + \epsilon_t \tag{1}$$

where ϵ_t is the error, μ is the mean of the time series, and ϕ_h is the parameter.

On the other hand, moving average (MA) models predict future values by incorporating past forecast errors. An MA(q) model of order q is represented by

$$Y_t = \sum_{k=1}^{q} \theta_k e_{t-k} + \mu + \epsilon_t \tag{2}$$

where θ_k are the parameters, and ε_t and e_{t-k} are the white noise processes.

The ARMA (p, q) model combines both AR and MA components to capture data dependencies. It is formulated as:

$$Y_t = \sum_{h=1}^{p} \phi_h Y_{t-h} + \sum_{k=1}^{q} \theta_k e_{t-k} + \mu + \epsilon_t \tag{3}$$

Finally, the Autoregressive Integrated Moving Average (ARIMA) (p, d, q) model extends the Autoregressive Moving Average (ARMA) by introducing differencing to make nonstationary data stationary. The model is described as follows.

$$W_t = \sum_{h=1}^{p} \phi_h Y_{t-h} + \sum_{k=1}^{q} \theta_k e_{t-k} + \mu + \epsilon_t \tag{4}$$

where $W_t = \Delta^d Y_t = (1-B)^d Y_t$, d represents the differencing degree, and B is the backshift operator.

Seasonal Autoregressive Integrated Moving Average Models

Seasonal Autoregressive Integrated Moving Average (SARIMA) models are expansions of regular Autoregressive Integrated Moving Average (ARIMA) models that can capture seasonal variations in time-series data. These models help work with data that display seasonality or recurrent patterns at regular intervals such as monthly, quarterly, or annual cycles. SARIMA (p, d, q) (P, D, Q)[s], where (p, d, q) denotes the non-seasonal ARIMA components (P, D, Q)[s] denotes the seasonal ARIMA components, and is the acronym for the seasonal

ARIMA model. For example, an ARIMA(1,1,1)(1,1,1)[4] model (without a constant) can be used for quarterly data(m=4) as follows:

$$(1-\phi_1 B)(1-\Phi_1 B^4)y_t = (1+\theta_1 B)(1+\Theta_1 B^4)\epsilon_t$$

While the SARIMA(1,1,2)(0,1,1) model requires the estimation of 17 parameters (12 for seasonality and 5 for autoregressive and moving average components), the seven years of data (89 observations) are sufficient for model fitting. Comparable models have been successfully applied to shorter time series in a similar study [20].

Various tests and models were utilized to assess the time series data. A unit root test, such as the Augmented Dickey-Fuller (ADF) test or Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, is employed to determine whether a time series is stationary or requires differencing. The Phillips-Perron test is an extension of the Dickey-Fuller test and tests the opposite hypothesis. To detect monotonic trends in the time series data, the Mann-Kendall test was used. The test statistic was computed as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}(x_i - x_j) \tag{6}$$

where S represents the Mann-Kendall test statistic, n is the number of observations, and $sgn(x_i - x_j)$ is the sign of the difference between the ith and jth observations. Positive values indicate an increasing trend, negative values indicate a decreasing trend, and zero indicates no trend [21].

Welch's t-test, a two-sample location test, was used to test for seasonal variations by comparing the means across different periods. The test statistic is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{7}$$

where $\bar{x}_i$ denotes the sample mean, S_{2i}^2 is the sample variance, and n_1 is the sample size for populations i=1,2.

The Generalized Linear Model (GLM) extends the linear model framework to accommodate response variables that follow distributions from the exponential family, such as normal, binomial, Poisson, and negative binomial distributions [17]. For Poisson regression, the probability distribution is given by

$$f(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{8}$$

where λ represents the expected number of occurrences and the log-transformed model is log(λi) = x'iβ. Parameters β and η are estimated using the maximum likelihood, with η indicating the degree of dispersion.

The Negative Binomial distribution is expressed as:

$$P(Y=y_i) = \frac{\Gamma(\theta+y_i)}{\Gamma(y_i+1)\Gamma(\theta)} \left(\frac{\lambda_i}{\lambda_i+\theta}\right)^{y_i} \left(1-\frac{\lambda_i}{\lambda_i+\theta}\right)^{\theta} \tag{9}$$

with the conditional mean E[yi/xi] = λi = exiβ and variance Var [yi/xi] = exiβ (1+η2exiβ).

Model diagnostics include the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to assess the trade-off between model fit and complexity [22]. Also, the Ljung-Box Q-statistic test was used to assess residual autocorrelation, ensuring that the model residuals are independent and free from significant autocorrelation.

$$Q_m = n(n+2)\sum_{k=1}^{m} \frac{r_k^2}{n-k} \tag{10}$$

where r_k^2 is the autocorrelation of the residuals at lag k; n is the number of residuals; and m is the number of time lags included in the test.

Forecast accuracy is measured by Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). RMSE is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2} \tag{11}$$

where y_t is the actual value, $\hat{y}_t$ is the forecasted value, and n is the sample size. MAPE is:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{A_t - F_t}{A_t}\right| \times 100\% \tag{12}$$

where A_t is the actual value, and F_t is the forecast value. Lower RMSE and MAPE values indicate better model performance.

Model Selection and Adequacy

We used a combination of statistical metrics and diagnostic checks to determine the best model. Various SARIMA models were evaluated, and the best model was considered the one with the lowest AIC, AICc, and BIC values, which implied an optimal balance between fit and complexity. The Ljung-Box Q-test was used to evaluate the seasonal trends and patterns in the data. Forecasting accuracy was assessed using MAPE and MASE, with lower values indicating better forecasting performance.

Statistical Analysis

The analysis was performed using the R-console version 4.3.1 statistical package. To analyze the time series, we applied Seasonal Autoregressive Integrated Moving Average (SARIMA) models to determine trends and seasonal trends in the incidence of scabies. The parameters of the model were selected using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), with lower values indicating improved model fit. The Corrected Akaike Information Criterion (AICc) was also used to account for small sample sizes, providing a more accurate measure of model fit. Residual diagnostics were conducted with the Ljung-Box Q-test for independence and absence of autocorrelation.

To model the count data, we compared Poisson and Negative Binomial regression models. We used the Negative Binomial model since it can accommodate overdispersion, as favored by a lower AIC value. Model performance was quantified by Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE), with lower values indicating more accurate forecasting. Statistical tests were conducted at a 0.05 significance level in all instances.

The stationarity of the scabies time-series data was assessed using Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. The test revealed that the original data were not stationary at a 5% significance level. To address this, first-order differencing was applied to remove the trend component by subtracting each observation from the previous one, seasonal differencing was also employed resulting in a stationary and non-seasonal series.

Poisson and Negative Binomial regression models were used to analyze the incidence of scabies. The Poisson model assumes that the mean and variance of count data are equal. However, this assumption often leads to overdispersion when the variance exceeds the mean, which is a common characteristic of real-world data. Overdispersion can result in a Poisson model underestimating the standard errors, leading to misleading conclusions [23]. To address overdispersion, we employed a Negative Binomial regression model, which includes an additional dispersion parameter to accommodate the variability in the data better.

Data Splitting and Model Validation

The dataset comprising 89 monthly observations from January 2016 to May 2023 was chronologically divided into training and test sets to ensure robust model validation. The training set comprised 71 observations (January 2016 to November 2021), representing approximately 80% of the total dataset, while the test set included the final

18 observations (December 2021 to May 2023), representing 20% of the data. This chronological split preserves the temporal structure essential for time series validation and allows assessment of the model's forecasting performance on unseen future data. The SARIMA model was fitted using the training data, and forecasting accuracy was subsequently evaluated on the test set using established metrics, including Mean Absolute Percentage Error (MAPE), Mean Absolute Scaled Error (MASE), ACF1, and Theil's U statistic.

Forecast Accuracy Metrics

Autocorrelation Function at lag 1 (ACF1) measures the correlation between consecutive forecast errors, with values closer to zero indicating better model performance and absence of systematic bias in residuals. Theil's U statistic is a relative accuracy measure comparing forecast performance to a naive benchmark, where values less than 1 indicate superior forecasting performance compared to a simple random walk model. Mean Absolute Percentage Error (MAPE) provides a percentage measure of forecasting accuracy, while Mean Absolute Scaled Error (MASE) offers a scale-independent measure that compares forecast accuracy against a seasonal naive forecast.

Results

The dataset comprises 89 observations of monthly scabies cases, ranging from a minimum of 99 to a maximum of 412 cases, with an average of 259.10 cases and a standard deviation of 75.79.

Exploratory Data Analysis and Visualization

Initial data exploration revealed that the scabies case data are slightly right-skewed (skewness = 0.23) and have heavier tails (kurtosis = 2.22) than a normal distribution. Figure 1 reveals several key insights: there is a consistent rise in scabies cases within the region over time, as shown by the upward slope in the trend component. From Figure 1, a distinct seasonal trend is evident, with scabies reaching its highest levels during the dry seasons (typically June-August) and its lowest levels during the rainy seasons (usually September-November). The remainder of the component shows fluctuations in the data that are not explained by trends or seasonal patterns, which appear to be relatively small compared with the overall signal.

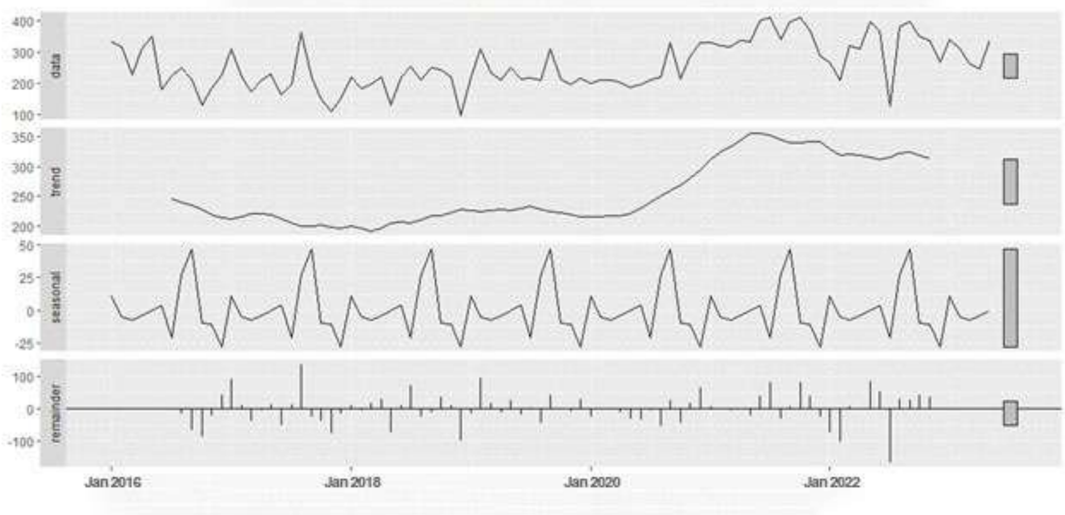

Figure 1. A decomposed plot of the scabies time series data.

The Mann-Kendall test confirmed a significant upward trend (p = 0.006), while Welch's test demonstrated significant seasonality (p = 0.009) in the scabies case data (Table 1),

Table 1. Mann-Kendall and Welch Tests for Trend and Seasonality.

Test	Test Statistic	P-Value	Conclusion
Mann-Kendall	4.32	0.006	Significant upward trend
Welch	3.75	0.009	Significant seasonality present

Stationarity Analysis

Stationarity analysis revealed that the original time series data were non-stationary, requiring first-order differencing to achieve stationarity (Table 2).

Table 2. Tests for stationarity of scabies cases.

Test	Order of Differencing	p-value	Conclusion
ADF	Original Data	0.062	Not Stationary
	After First Differencing	0.01	Stationary
KPSS	Original Data	0.01	Not Stationary
	After First Differencing	0.1	Stationary
Welch	Original Data	0.01	Seasonal
	After Differencing	0.92	Not Seasonal

Model Selection and Adequacy

Table 3 presents a comparison of various SARIMA models based on their Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc), and Bayesian Information Criterion (BIC) scores. Among the models evaluated, SARIMA (1,1,2)(0,1,1) recorded the lowest AIC (866.79) and BIC (878.44) values, indicating that it provides the best balance between model fit and complexity. This suggests that the SARIMA(1,1,2)(0,1,1) model is the most appropriate for forecasting scabies incidence in the study setting

Table 3. Comparison of SARIMA models based on AIC, AICc, and BIC.

SARIMA Model	AIC	AICc	BIC
SARIMA(0,1,3)(0,0,1)	977.06	977.79	989.44
SARIMA(1,1,2)(1,0,0)	977.84	978.58	990.23
SARIMA(1,1,1)(0,0,1)	981.02	981.50	990.92
SARIMA(1,1,3)(0,0,1)	979.00	980.04	993.87
SARIMA(2,1,2)(1,0,0)	979.48	980.52	994.35
SARIMA(0,1,3)(0,1,0)	884.39	884.95	893.71
SARIMA(1,1,2)(0,1,1)	866.79	867.65	878.44
SARIMA(0,1,3)(1,1,0)	872.96	873.82	884.62
SARIMA(0,1,3)(1,1,1)	869.20	870.42	883.19

AIC: Akaike Information Criterion;
AICc: Corrected Akaike Information Criterion;
BIC: Bayesian Information Criterion

Model Adequacy, Residuals Diagnostic, Parameter Estimates, and Forecast Accuracy

The SARIMA(1,1,2)(0,1,1) model was evaluated using various tests of model adequacy, residual diagnostics, and forecast accuracy. The parameter estimates are also presented below, confirming the adequacy of the SARIMA(1,1,2)(0,1,1) model for forecasting scabies cases in the Ashanti Region, as indicated in Table 4.

Table 4. Model adequacy, residuals diagnostic, parameter estimates, and forecast accuracy for SARIMA(1,1,2)(0,1,1).

Category	Metric/Test	Value	Conclusion
Model Adequacy	Ljung-Box Test Statistic	7.55	Model is appropriate
	Ljung-Box P-Value	0.9113	
Residuals Diagnostic	Lilliefors Test Statistic	0.093	Residuals normally distributed
	Lilliefors P-Value	0.062	
Parameter Estimates	AR1 Coefficient	-0.76	Standard Error: 0.08
	MA1 Coefficient	0.26	Standard Error: 0.13
	MA2 Coefficient	-0.74	Standard Error: 0.11
	SMA1 Coefficient	-0.67	Standard Error: 0.15
Forecast Accuracy	Training Set MAPE	18.89	
	Test Set MAPE	17.65	
	Training Set MASE	0.60	
	Test Set MASE	0.68	
	Test Set ACF1	0.06	
	Test Set Theil's U	0.87	

Extended Forecasted Values

The SARIMA (1,1,2) (0,1,1) model was trained on the full dataset (January 2016 to May 2023) and used to generate out-of-sample forecasts for June 2023 to May 2025, as shown in Table 5. The extended forecast values for scabies cases from June 2023 to May 2025 using the SARIMA(1,1,2)(0,1,1) model, along with 95% confidence intervals, are presented below. Based on the forecast, after an initial dip, scabies cases are projected to increase over time with steady fluctuations. The extended forecast indicates a similar pattern to the previous forecast: a dip in cases in mid-2023 followed by a steady rise through mid-2025, with seasonal fluctuations.

Table 5. Extended forecasted values for scabies cases.

Month	Forecast	Lower (95%)	Upper (95%)
Jun 2023	315	187.3441	443.5746
Jul 2023	259	117.5761	401.6524
Aug 2023	326	183.2671	469.9287
Sep 2023	373	221.0351	525.4804
Oct 2023	318	164.3329	473.5125
Nov 2023	317	156.3055	478.5499
Dec 2023	276	112.0051	440.3227
Jan 2024	315	145.6561	484.5456
Feb 2024	290	117.4257	463.0270
Mar 2024	290	113.1144	467.9087
Apr 2024	288	107.5599	469.3446
May 2024	334	149.4439	519.5338
Jun 2024	349	160.6445	538.6102
Jul 2024	293	114.3224	473.5419
Aug 2024	364	176.3891	553.3555
Sep 2024	415	218.9127	613.0401
Oct 2024	355	156.5973	555.1301
Nov 2024	354	147.2683	561.1240
Dec 2024	307	99.4834	516.2697
Jan 2025	352	137.9125	566.1266
Feb 2025	320	104.3429	536.4940
Mar 2025	320	100.4312	541.0307
Apr 2025	318	94.2989	542.6500
May 2025	369	134.1276	604.9608

Regression Model Comparison: Poisson vs. Negative Binomial

The Negative Binomial model demonstrated superior performance with significantly lower AIC and BIC values (1039.6 vs. 2496.4 and 1071.9 vs. 2526.2, respectively), indicating a better fit for the data. The Negative Binomial model results revealed significant overdispersion, evidenced by a high dispersion parameter of 23.43 and a very low p-value (5.34e-14). Significant increases in scabies cases were observed in June and July, whereas November and December saw significant decreases, highlighting seasonal patterns and variability in the incidence of scabies.

Table 6. Negative binomial regression model output.

Month	Estimate	Pr(>|z|)
January	-	-
February	-0.064807	0.650
March	-0.134621	0.346
April	-0.081201	0.570
May	-0.005844	0.967
June	0.111018	0.043
July	0.173004	0.024
August	0.013673	0.926
September	0.084853	0.565
October	0.124852	0.059
November	-0.130672	0.037
December	-0.208334	0.015

Discussion

This study demonstrated a significant upward trend in scabies incidence in the Ashanti Region of Ghana, with distinct seasonal patterns peaking during the dry season. The SARIMA(1,1,2)(0,1,1) model provided accurate forecasts, while the Negative Binomial regression model identified significant seasonal variations.. The SARIMA(1,1,2)(0,1,1)[12] model has emerged as the most suitable for forecasting monthly scabies incidence. This model effectively captured the seasonal patterns observed in the data, with peak incidence occurring during the wet season. The time-series analysis revealed a consistent increase in scabies cases over the study period, highlighting the growing public health concern. The distinct seasonal trend, with cases peaking during the dry season and reaching their lowest points during the rainy season, underscores the importance of seasonal preparedness in healthcare facilities. The Negative Binomial regression model outperformed the Poisson model, indicating significant overdispersion in the data. This model identified significant seasonal variations, with a notable increase in the number of cases from June to October. July emerged as the month with the highest incidence, providing crucial information for resource allocation and public health planning. Our findings align with recent studies that have employed SARIMA models to analyze scabies incidence. For instance, Perez et al. [16] reported similar seasonal patterns in scabies infections among wild ibex, while Ghiebi et al. [15] demonstrated the applicability of SARIMA models to vector-borne diseases, including scabies and also align with the literature on scabies in tropical regions, which often reports a higher prevalence in crowded living conditions and among vulnerable populations such as children and the elderly. The seasonal pattern observed may be related to factors such as changes in hygiene practices, population movements, or environmental conditions that favor mite survival and transmission. The forecasted increase in scabies incidence underscores the need for sustained public health interventions. The seasonal patterns identified in this study provide valuable insights for resource allocation, particularly during peak months (June to October). Our study is one of the first to apply SARIMA and Negative Binomial models to scabies incidence in a resource-limited setting. The findings align with previous studies that have reported seasonal patterns in scabies infections, such as Perez et al. [16], who observed similar trends in wild ibex populations.

This study has several limitations. First, the data were obtained from a single hospital, which may limit the generalizability of the findings. Second, the time series was relatively short (89 months), which may affect the stability of the model parameters. Future studies should aim to include data from multiple sources and over longer periods. The findings of this study have important implications for public health policy and practice. By anticipating periods of high incidence, healthcare facilities can better prepare for increased demand and implement targeted interventions. Future research should explore the impact of environmental and socioeconomic factors on scabies transmission. While this study did not explicitly analyze the influence of sex or gender on scabies incidence, previous research has shown that children and the elderly are more vulnerable to scabies infections. Future studies should explore the role of demographic factors in scabies transmission

Clinical and Public Health Implications

The findings from this study have several important implications for healthcare planning and scabies management in the Ashanti Region and similar settings. Healthcare facilities should enhance their preparedness for scabies cases during peak transmission periods, particularly from June to October, by ensuring adequate medication supplies and staffing allocation. Public health authorities should implement targeted awareness campaigns focusing on prevention strategies, symptom recognition, and treatment adherence, timed to precede and coincide with peak seasons.

The establishment of robust surveillance systems utilizing the forecasting models developed in this study would facilitate early detection of potential outbreaks and enable timely intervention measures. Resource allocation strategies should incorporate the seasonal patterns identified, with increased healthcare capacity during predicted high-incidence months. Community-based preventive measures targeting overcrowded living conditions, improved hygiene practices, and enhanced access to clean water represent critical components of comprehensive scabies control programs.

Healthcare worker training programs should emphasize rapid diagnosis and appropriate treatment protocols, particularly in preparation for seasonal peaks. Integration of these forecasting models into existing health information systems could support evidence-based decision making and optimize healthcare resource utilization in resource-limited settings like the study location.

Conclusion

This study demonstrated the effectiveness of combining time-series analysis and generalized linear models in understanding and predicting the incidence of scabies. The SARIMA(1,1,2)(0,1,1) model provides a reliable tool for forecasting monthly cases, whereas the Negative Binomial regression model offers insights into seasonal variations. This research highlights a concerning upward trend in scabies cases in the Ashanti Region, with clear seasonal patterns. The highest incidence occurred during the wet season, particularly from June to October, with July showing the most significant increase in incidence. Conversely, November and December showed a significant decrease in the cases. These findings have important implications in public health planning and resource allocation. By anticipating periods of high incidence, healthcare facilities can better prepare for increased demand and implement targeted interventions measures. The forecasted increase in scabies incidence underscores the need for sustained public health interventions. The seasonal patterns identified in this study provide valuable insights for resource allocation, particularly during peak months (June to October).

List of Abbreviations: SARIMA – Seasonal Autoregressive Integrated Moving Average; AIC – Akaike Information Criterion; BIC - Bayesian Information Criterion; MAPE – Mean Absolute Percentage Error; MASE – Mean Absolute Scaled Error; ARIMA – Autoregressive Integrated Moving Average; ARMA – Autoregressive Moving Average; ADF – Augmented Dickey-Fuller; KPSS – Kwiatkowski-Phillips-Schmidt-Shin; GLM – Generalized Linear Model

Author Contributions: M.T. and M.A.N. reviewed the study's proposal. E.O.O. and O.S. defined the research aim and design of the experiment. M.A.O., W.A.P., and S.A.H. participated in the design of the study and performed statistical analysis. P.S.A. and E.A.A. coordinated and helped to draft the manuscript. All the authors have read and approved the final manuscript.

Funding: This research received no funding.

Ethics Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study were obtained from Kokofu Hospital, Ashanti Region, Ghana. The data were not publicly available because of patient confidentiality agreements. However, researchers interested in accessing the data may contact the hospital administration for permission.

Conflict of Interest: The authors declare no conflict of interest.

References

1. Abolhasani R, Araghi F, Tabary M, Aryannejad A, Mashinchi B, Robati RM. The impact of air pollution on skin and related disorders: a comprehensive review. Dermatol Ther. 2021;34(2):e14840. https://doi.org/10.1111/dth.14840.
2. Romani L, Steer AC, Whitfeld MJ, Kaldor JM. Prevalence of scabies and impetigo worldwide: a systematic review. Lancet Infect Dis. 2015;15(8):960-7. https://doi.org/10.1016/S1473-3099(15)00132-2.
3. Welch E, Romani L, Whitfeld MJ. Recent advances in understanding and treating scabies. Faculty Reviews. 2021;10:28. https://doi.org/10.12703/r/10-28.
4. Centers for Disease Control and Prevention. Parasites – Scabies. Atlanta, GA: CDC; 2020 [cited 2025 Aug 19]. Available from: https://www.cdc.gov/parasites/scabies
5. Schneider S, Wu J, Tizek L, Ziehfreund S, Zink A. Prevalence of scabies worldwide—An updated systematic literature review in 2022. J Eur Acad Dermatol Venereol. 2023 ;37(9):1749-57. https://doi.org/10.1111/jdv.19167.
6. Alabbadi A, Alalwi S, Alkhalifah R, Alfaraj D, Albreiki S, Rubaian NB, Alsagheer M. Scabies-associated leukocytoclastic vasculitis: A case report and review of literature. Cureus. 2021;13(12):e20725. https://doi.org/10.7759/cureus.20725.
7. Sunderkötter C, Wohlrab J, Hamm H. Scabies: epidemiology, diagnosis, and treatment. Deutsches Ärzteblatt International. 2021;118(41):695. https://doi.org/10.3238/arztebl.m2021.0296.
8. Amoako YA, Phillips RO, Arthur J, Abugri MA, Akowuah E, Amoako KO, et al. A scabies outbreak in the North East Region of Ghana: The necessity for prompt intervention. PLoS Negl Trop Dis. 2020;14(12):e0008902. https://doi.org/10.1371/journal.pntd.0008902.
9. Birjandi MM, Oroei M, Emadi SN, Peyvandi AA, Anang AK. Scabies among high school students in Accra, Ghana: Risk factors and health literacy. Iranian Red Crescent Medical Journal. 2019;21(8):1-8. https://doi.org/10.5812/ircmj.92510.
10. Kaburi BB, Ameme DK, Adu-Asumah G, Dadzie D, Tender EK, Addeh SV, et al. Outbreak of scabies among preschool children, Accra, Ghana, 2017. BMC Public Health. 2019;19(1):746. https://doi.org/10.1186/s12889-019-7085-6.
11. Cox V, Dowden M, Williams C, Engelman D, Yan J, Ralph A, et al. P25 Considering the burden of scabies alongside a mass drug administration treatment programme in an indigenous community in northern Australia. Br J Dermatol. 2023;188(Supplement_4):ljad113.053. https://doi.org/10.1093/bjd/ljad113.053.
12. Williams R. Review of Regression Models for Categorical Dependent Variables Using Stata, Second Edition, by Long and Freese. Stata J. 2006;6(2):273-8. https://doi.org/10.1177/1536867X0600600208.
13. El-Moamly AA. Scabies as a part of the World Health Organization roadmap for neglected tropical diseases 2021–2030: what we know and what we need to do for global control. Trop Med Health. 2021;49(1):64. https://doi.org/10.1186/s41182-021-00348-6.
14. Engelman D, Marks M, Steer AC, Beshah A, Biswas G, Chosidow O, et al. A framework for scabies control. PLoS Negl Trop Dis. 2021;15(9):e0009661. https://doi.org/10.1371/journal.pntd.0009661.
15. Gheibi Z, Boroomand M, Soltani A. Comparing the Trends of Vector‐Borne Diseases (VBDs) before and after the COVID‐19 Pandemic and Their Spatial Distribution in Southern Iran. Journal of Tropical Medicine. 2023;2023(1):7697421. https://doi.org/10.1155/2023/7697421.
16. Perez JM, López-Montoya AJ, Cano-Manuel FJ, Soriguer RC, Fandos P, Granados JE. Seasonal Dynamics of Scabies Infections in Wild Ibex Populations in Southern Spain. J Wildl Manag. 2022;86(5):e22224. https://doi.org/10.1002/jwmg.22224.
17. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46(1):348-55. https://doi.org/10.1093/ije/dyw098.

18. Pewsey A, García-Portugués E. Recent advances in directional statistics. Test. 2021;30(1):1-58. https://doi.org/10.1007/s11749-021-00759-x.
19. Phillips PCB, Shi S, Yu J. Testing for unit roots with nonstationary volatility and abrupt structural change. Journal of Econometrics 2023;235(2):1086-110. https://doi.org/10.1016/j.jeconom.2023.02.007
20. Cavaliere G, Taylor AMR. Testing for unit roots in time series models with non-stationary volatility. Journal of Econometrics. 2007;140(2):919-47. http://dx.doi.org/10.1016/j.jeconom.2006.07.019.
21. Wong H, Ip WC, Zhang R, Xia J. Non-parametric time series models for hydrological forecasting. J Hydrol. 2007;332(3–4):337-47. https://doi.org/10.1016/j.jhydrol.2006.07.013.
22. Zhang J, Yang Y, Ding J. Information criteria for model selection. Wiley Interdisciplinary Reviews: Computational Statistics. 2023;15(5):e1607. https://doi.org/10.1002/wics.1607.
23. Lee W, Kim J, Lee D. Revisiting the analysis pipeline for overdispersed Poisson and binomial data. J Appl Stat. 2022;50(7):1455-1476. https://doi.org/10.1080/02664763.2022.2026897.