Rhiannon Loster
Department of Mathematics and Statistics
University of Guelph

Lyme disease spread in the United States – Modeling and forecasting cases from 2022 to 2024

Keywords: Lyme disease, United States, SARIMA model, STL analysis, Forecasting, Temperature.

Abstract

Monthly Lyme disease cases in the United States was investigated and a model was fit to forecast Lyme disease cases from January 2022 to December 2024. Temperature was assessed as a possible covariate for Lyme disease cases at the US national level. Finally, yearly Lyme disease cases for each state or locality in the US was investigated and fit to linear regression models to forecast yearly case totals per state/locality from 2022 to 2024. There was a seasonal change in US total Lyme disease cases, where a rise of one degree Celsius showed an increase in about one Lyme disease case. However, there was no evidence of an increasing or decreasing trend for US total cases, and seasonal peaks remained about 3,000 cases each year. Also, there was a shift from increasing to decreasing trends in yearly Lyme disease cases per state/locality when assessing trends over the last 13 years or over the last 5 years; showing a more recent decreasing trend for many states, especially those in southern US.

Introduction

Vector-borne diseases that are illnesses that are transmitted by vectors, which include mosquitoes, ticks, and fleas. These vectors can carry infective pathogens such as viruses, bacteria and protozoa which is transferred from one host (carrier) to another (Beard et al, 2016). Lyme disease or Lyme borreliosis is the most common vector-borne disease in the United States and is considered endemic in certain parts of the US, including the northeastern and north-central states (CDC, 2022; Skar & Simonsen, 2023). The main states where Lyme disease is endemic are Connecticut, Delaware, Maine, Maryland, Massachusetts, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, Virginia, and Wisconsin (Skar & Simonsen, 2023). Lyme disease is also among the most frequent tick-borne disease worldwide (Skar & Simonsen, 2023). There are about 30,000 Lyme disease cases reported to the CDC, through state health departments and the District of Columbia surveillance, every year (CDC, 2022). The CDC has stated that this is not reflective of all the cases of Lyme disease that are diagnosed in the US, and recent estimates of the number of people who get Lyme disease in the US to be approximately 476,000 a year (CDC, 2022).

Lyme disease is typically caused by the bacterial spirochete *Borrelia burgdorferi* and less commonly by *Borrelia mayonii* in the US (CDC, 2022; Skar & Simonsen, 2023). In Eurasia it is typically caused by both *Borrelia afzelii* and *Borrelia garinii* (Skar & Simonsen, 2023). *B.*

burgdorferi is transmitted by *Ixodes* genus tick and the most common for transmission is Ixodes scapularis, also known as the blacklegged tick (CDC, 2022; Skar & Simonsen, 2023). Host animals for the blacklegged tick include deer, rodents, birds, etc. and certain host animals, such as the white-footed mouse, harbour *B. burgdorferi* (EPA, 2023). When a blacklegged tick bites an infected host animal, it can transmit *B. burgdorferi* by biting another host animal or human (EPA, 2023).

Lyme disease presents itself in three stages; (1) early localized, (2) early disseminated, and (3) late stages (Skar & Simonsen, 2023). The early localized stage is characterized by an expanding red ring like rash called Erythema migrans at the site of the tick bite (Skar & Simonsen, 2023). Other symptoms at this stage include flu-like symptoms, headache, fever, malaise, and pain in the muscles and joints (Skar & Simonsen, 2023). Only 20 percent of patients develop into the early disseminated phase before they are treated, and it is characterized by multiple Erythema migrans lesions (Skar & Simonsen, 2023). Other symptoms of this stage include swelling lymph nodes, muscle and joint pain, palsies of the cranial nerves, ophthalmic conditions, lymphocytic meningitis, and heart conditions like inflammation of the heart and conduction abnormalities (Skar & Simonsen, 2023). Lastly, the late stage is characterized by arthritis in large joints, such as knees (Skar & Simonsen, 2023). The early localized phase is when the majority of patients are treated (Skar & Simonsen, 2023).

All stages of Lyme disease are currently treated with antibiotics (CDC, 2022; Rebman & Aucott, 2020). However, recent studies suggest that a subset of treated patients can have long term recurring symptoms and are diagnosed with post-treatment Lyme disease syndrome (Rebman & Aucott, 2020). Some of these long-term symptoms include fatigue, chronic muscle and joint pain, and cognitive issues (Rebman & Aucott, 2020). These symptoms present similarly to patients with myalgic encephalomyelitis and long Covid (Rebman & Aucott, 2020; Komaroff & Lipkin, 2023). Currently, prevention of Lyme disease remains only with non-pharmaceutical interventions, such as insect repellent, removing ticks promptly, pesticides, avoiding tick habitat like long grassy areas, and wearing long clothing (CDC, 2022).

Most tick bites and therefore Lyme disease infections occur during the late spring, summer, and early fall (Skar & Simonsen, 2023). This is typically when blacklegged ticks are most active because the temperature has risen above 45 degrees Fahrenheit, or about 7 degrees Celsius (EPA, 2023). Blacklegged ticks also thrive more in areas with at least 85 percent humidity, putting warmer humid areas more at risk for Lyme disease (EPA, 2023). Climate factors, such as temperature extremes and changes in precipitation patterns, can affect the seasonality, distribution, and prevalence of vector-borne diseases, including Lyme disease (Beard et al, 2016). Warming climates from climate change has increased the range of suitable blacklegged tick habitat, which has led to a generally northward expansion of these ticks (Beard et al, 2016, EPA, 2023). Climate change has also led to warmer temperatures in the shoulder seasons, increasing the length of the season (Beard et al, 2016). Climate change induced increases in seasonal activity and growing geographic range of blacklegged ticks will further increase the risk of human exposure to Lyme disease, as well as other tickborne diseases (Beard et al, 2016; CDC, 2022).

As linear regression models assume that the data is independent and no autocorrelation is present, the Durbin-Watson test statistic is used to assess any temporal autocorrelation in residuals resulting from a linear regression model that is fit to time series data (Chatfield, 1996, p.62/63). A two-sided test is generally recommended. A test statistic value between zero and two indicates a positive autocorrelation, a value between two and four indicates a negative autocorrelation, with stronger autocorrelation being closer to zero or four, respectively. Values close to two are considered to have weak autocorrelation.

A seasonal time series decomposition (STL) analysis can be used to determine if a data set has seasonal and/or trend component(s). The STL analysis is based on locally estimated smoothing splines (LOESS). An STL algorithm will fit LOESS smoothers depending on the trend and season "windows" defined. The information from this analysis can be used when fitting a Seasonal Autoregressive Integrated Moving Average (SARIMA) model to time series data.

Autoregressive Integrated Moving Average (ARIMA) time series models removes the assumption that there are no trend effects in time series data, by integrating and removing these trends from the time series (Box and Ljung, 1970). A seasonal ARIMA (SARIMA) model allows for both the seasonal effects and the trend to be integrated and removed from the time series. The fit of the model can be assessed using the Akaike Information Criterion (AIC), which indicates a better fit when the AIC is smaller. The AIC is preferable to the Bayesian Information Criterion (BIC) when using the model to forecast values (Shmueli, 2010).

A few techniques to check the fit of a model is to plot the auto-covariance/correlation-function (ACF), the partial auto-correlation function (PACF), and to perform a Ljung-Box test. The ACF and PACF are used to check for any remaining correlation in the residuals of the model. A Ljung-Box test can check for lack-of-fit in a model and checks for any remaining autocorrelation in the residuals at a variety of time lags (Ljung & Box, 1978). The Durbin-Watson test only detects first-order autocorrelation.

The goal of this study was to fit a model to forecast the number of Lyme disease cases in the United States for the following three years, 2022 to 2024, and to assess Lyme disease trends per state or locality in the US. To attain this goal, the specific objectives were to (i) explore the US total Lyme disease data, (ii) determine whether there was a trend and/or seasonal component to this data, (iii) fit a model to the data to forecast from January 2022 to December 2024, (iv) investigate the effect of temperature as a covariate, and (v) investigate and forecast Lyme disease data per state/locality.

Materials and Methods

Lyme disease monthly reported case data for the United States was retrieved from the CDC Lyme Disease surveillance website (CDC data, 2022). The data consisted of monthly total cases from 2008 to 2021, however, Lyme disease has been a nationally notifiable disease in the US since 1991 (CDC data, 2022). Yearly Lyme disease total case numbers for each state or locality data was also obtained from the CDC Lyme Disease surveillance website (CDC data, 2022). Case numbers are collected by state and local health departments, specific to each department's surveillance practices, and these reports are shared with the CDC (CDC data, 2022). Lastly, monthly temperature data for the US was retrieved from the National Centers for the Environmental

Information website (NOAA, 2023). The data was retrieved for the same available range as the Lyme disease case data, from 2008 to 2021.

First, the US monthly Lyme disease cases were investigated. Descriptive statistics were found for the data, and the distribution of the data was assessed using both a boxplot and histogram of the case data and the logarithm of the case data. A Durbin-Watson test was conducted using the *dwtest* function between Lyme disease case data and the year-month time to assess the presence of correlation within the data (Chatfield, 1996). Next, the monthly Lyme disease cases was plotted and smooth splines were added using the *smooth.spline* function with different spar values to visually investigate any seasonal change or trend within the data. This was also used to investigate the logarithmic monthly case data.

For both the monthly Lyme disease case data and the logarithmic month case data, an STL analysis was performed to explore possible trends and seasonal components of the data. This was done using the *stl* function and the STL analysis for each was plotted to assess the choices for *s.window* and *t.window* values. For both STL analyses the auto-covariance/correlation-function (ACF) was plotted using the acf function to check for any autocorrelation of the remainders.

Using the interpreted d and D values from the STL exploratory, SARIMA models were fitted to both sets of data using the *auto.arima* function. Each model was assessed by comparing the Akaike Information Criterion (AIC), mean absolute error (MAE), and the root mean square error (RMSE), to determine which one was more accurate. ACF, partial-ACF (PACF), and QQ plots were used to check the fit of the better performing model, using the *acf*, *pacf*, and *qqplot* functions, respectively. A Ljung-Box test was also performed to check for autocorrelation of the residuals and the *tsdiag* function was used to check the p-values for this test at a variety of lag values (Ljung & Box, 1978). The SARIMA model was then used to make a projection of Lyme disease cases for the US for the next 36 months, from January 2022 to December 2024 with a 95 percent confidence interval.

Next, the monthly US temperature data was used to investigate the effects of temperature on Lyme disease cases. First a scatter plot with a smooth spline was used to visually assess any possible correlation between Lyme disease case numbers and temperature change. A correlation test, using the Spearman method was done to further investigate any correlation. The time series data for temperature and the logarithmic time series data for Lyme disease cases was further visualized and this was fitted to an SARIMA model using the *auto.arima* function. The ACF, PACF, and QQ plots were used to check the fit of this model. A function by Rolf Turner, 2009 was used to find the coefficients and their p-values for the SARIMA model, to determine the influence temperature has on the number of Lyme disease cases in the US.

Lastly, the yearly Lyme disease data for each state/locality was investigated using simple linear regression. Descriptive statistics of the yearly Lyme disease cases from 2008 to 2021 were obtained for each location and tabulated. The data from each location was plotted and a smooth spline was used to visualize any trends over the last 13-year and 5-year periods. A linear regression using the *lm* function was done for each of these periods and for each location. These were plotted with each state/locality data to visually assess the trends. The trends of each location, determined by the slope found during the linear regression, were then represented on a map for the last 13 and 5-year periods, to visualize the longer trend and more recent Lyme disease trends for each location. The linear regression model for both the 13-year trend and the 5-year trend were then projected using the *forecast* function for each state from 2021 to 2024. Note that any forecasted value below zero was considered as zero cases.

All methods described above were done in RStudio (RStudio Team, 2020), and the *auto.arima*, forecast, and *dwtest* functions were found in the R library "forecast" and "lmtest" respectively (Hyndman et al, 2023; Zeilis & Hothorn, 2002).

Results

The Lyme disease monthly cases data for the United States ranged from 2008 to 2021, with observations each month, resulting in 168 observations. The median and mean reported cases were 1068 and about 1973 cases, respectively (variance = 4267662), with a range of 262 to 8310 cases. A boxplot and histogram of the data is shown in Figure 1. Both of these plots indicate non-Gaussian data, so a further boxplot and histogram of the logarithmic values of the data was plotted and is shown in Figure 2. These indicated that the logarithmic data had a Gaussian distribution. The Durbin-Watson test statistic was about 0.675 with a p-value of less than 2.2e-16, which indicated that there was some positive autocorrelation within the case data with reference to the year-month and that this test was significant. Going forward, both the logarithmic and regular case data were investigated. Figure 3 shows the original (top) and logarithmic (bottom) case data were plotted, and smooth splines were used to show possible trend and seasonal components to the data.

When completing an STL analysis for the original case data, a trend window value of 168 was chosen to fit the entire range of data and a seasonal window value of 13 was chosen as the exploratory had indicated a possible yearly or 12-month season. However, there was still indication of autocorrelation at lag = 1, and so the seasonal window value was changed to 7 to account for the autocorrelation in the residuals. Figure 4 shows a plot of the seasonal, trend, and residual components of this STL analysis. A similar STL analysis for the logarithmic case data, with a trend window value of 168 and seasonal window value of 13, gave the seasonal, trend, and residual component plots shown in Figure 5. Furthermore, a plot of the original case data and logarithmic case data with their corresponding found trend and seasonal components are shown in Figure 6 and 7, respectively. Lastly, the ACF of both analyses were plotted in Figure 8. Both analyses indicated that the trend component for the SARIMA model should be set to zero $(d = 0)$ and the seasonal component for the SARIMA model should also be set to one $(D = 1)$.

When fitting a SARIMA model to the original case data, the parameters found for the trend component (p, q) and the seasonal component (P, Q), were $(p, q) = (1, 1)$ and $(P, Q) = (2, 0)$. While the parameters found when fitting a SARIMA model to the logarithmic case data were $(p, q) = (2, 1)$ and $(P, Q) = (2, 0)$. The AIC for the first model was 2343.77, the MAE was about 220.18, and the RMSE was about 400. There was indication of lack-of-fit for the original case data SARIMA model in the QQ plot shown in Figure 9. For the second logarithmic case model, the AIC was -763.37, the MAE was about 0.1001, and the RMSE was about 0.129. As the SARIMA model fit to the logarithmic case data scored better on all model fit assessments indicated, it was chosen for analysis moving forward. Neither the ACF/PACF plots nor the QQ plot for the logarithmic case model, shown in Figure 10 and 11, respectively, indicated any lack-of-fit. Lastly, the Ljung-Box test was not significant at two times the cycle length (p-value = 0.1003), indicating that there was no lack-of-fit for the model.

The projected logarithmic Lyme disease cases using the fitted SARIMA model is shown in Figure 12. This projects the maximum Lyme disease cases for each year to be 2,852.574 cases (CI 95% = (1577.456, 5410.214)) in July 2022, 3,512.014 cases (CI 95% = (1822.28, 7168)) in July 2023, and 2,961.1 cases (CI 95% = (1393.695, 6804.66)) in July 2024.

During the 2008 to 2021 period, the median and mean temperature were 11.86 and 11.99 degrees Celsius, respectively (variance = 71.0736). The data ranged from -1.31 to 24.87 degrees Celsius. The scatter plot of cases versus temperature is shown in Figure 13 and is indicative of a possible relationship. The correlation test using the spearman method gave a correlation value of about 0.74 (CI 95% = (0.67, 0.80)), with a p-value of less than 2.2e-16, which also indicated some positive correlation between temperature and cases. The logarithmic case data time series object was plotted with the temperature time series object to further visualize the seasonal correlation and is shown in Figure 14. The SARIMA model that was fitted to the logarithmic case data with temperature indicated as a potential covariate. None of the ACF, PACF, or QQ plots indicated any lack of fit for this model, and the AIC value was -772.25, which is slightly less good fit compared to the logarithmic case SARIMA model discussed previously. The coefficient value for temperature in the SARIMA model was 0.0027869352 (p-value = 0.0032553908), which indicates that this effect is significant. As this gives the coefficient in logarithmic cases, it actually shows that for one-degree Celsius increase in temperature the Lyme disease cases would increase by about 1.002 Lyme disease cases.

The linear regressions of each state or locality in the US for the 13-year period (2008 to 2021) had 31 states/localities with a positive/increasing trend in Lyme disease cases, 19 states/localities with a negative/decreasing trend in Lyme disease cases, and one state/locality with no trend. The mean positive trend in cases was a slope equal to 20.86 cases per year (range = (0.068, 209.69)) and the mean negative trend in cases was a slope equal to -63.10 cases per year (range = (-444.78, -1.27e-14)). Figure 15 shows a map representation of the states'/localities' trends. The linear regressions of each state or locality in the US for the 5-year period (2017 to 2021) had 15 states/localities with a positive/increasing trend in Lyme disease cases, 35 states/localities with a negative/decreasing trend in Lyme disease cases, and again, one state/locality with no trend. The mean positive trend in cases was a slope equal to 42.81 cases per year (range = (0.9, 268.2)) and the mean negative trend in cases was a slope equal to -166.73 cases per year (range = (-2487.4, -0.2)). Similarly, Figure 16 shows a map representation of the states'/localities' trends. There was a shift in state/locality trends when investigating the regression from 2008 to 2021 and the regression from 2017 to 2021. The more recent trends (5-year regression) showed far more short term decreases in Lyme disease for more states, while the longer trend (13-year regression) showed more increase in Lyme disease. The states/localities that switched from a positive trend in the 13-year regression to a negative trend in the 5-year regression, included Alabama, California, Colorado, Florida, Kansas, Louisiana, Maine, Mississippi, Missouri, Nebraska, Nevada, North Dakota, Oregon, Pennsylvania, Rhode Island, Utah, Vermont. The only state that switched from a negative to a positive trend was Georgia, and the rest remained unchanged.

The forecast for each state/locality continued with the trends shown in the 13-year or 5-year regression period. Each state's/locality's 2024 forecasts are tabulated for the 13-year regression model and the 5-year regression model in Table 1 and 2, respectively. The 13-year regression forecast of 2024 has a range from 0 to about 8,764 cases, with the highest being Pennsylvania, while the 5-year regression forecast of 2023 has a range from 0 to about 2,353 cases, with the highest being West Virginia. The 2024 forecasted total Lyme disease cases for each state/locality showed a similar shift in trend as discussed above, with many more states reaching zero cases by 2024.

Discussion

The fitted SARIMA model showed no continuing trend in Lyme disease cases in the United States throughout the projected 2022 to 2024 period, with the seasonal peaks, but showed a continuing seasonal component, with Lyme disease case numbers typically hitting a peak around 3,000 cases in July and reaching a low of around 200 cases in February each of the three projected seasons. With no visible trend seen in the STL analysis, but an obvious seasonal component within the analysis on the logarithmic case data from January 2008 to the end of 2021, it was not surprising to see a continuing seasonal component within the model projection.

As there was a consistent seasonal component in Lyme disease case data, temperature as a possible covariate for the data was investigated and a significant spearman correlation value of 0.74 was found between temperature change and the logarithmic US Lyme disease cases. It was also found that an increase in one degree Celsius in temperature would increase the number of Lyme disease cases by about one case. The effect of temperature as a covariate for Lyme disease cases was not surprising as the STL analysis showed an obvious seasonal trend. Also, since blacklegged ticks are known to be more active about seven degrees Celsius, it is reasonable that an increase in temperature would affect the likelihood of humans being bit by these ticks.

Total Lyme disease cases increasing in the US due to climate change, but this is not the case in this study there was no found trend, whether increasing or decreasing (EPA, 2023). However, this study highlighted the effect temperature has on Lyme disease cases as a covariate and the projection does not take an increase in overall temperature into account. Therefore, there could be an increasing trend in cases if temperatures continue to rise. Future research could use projected temperature forecasts to include within a Lyme disease case forecast for possibly more accurate predictions.

The large shift in trends from increasing Lyme disease cases in the longer 13-year regression model to decreasing Lyme disease cases in the 5-year regression model for some states, is indicative of positive change for these states/localities. Part of this shift could be due to the shift in tick hosts in the southern states (Ginsberg et al, 2021). This shift from mammalian to reptilian hosts, which are not as efficient reservoirs for bacterial spirochetes, such as *Borrelia burgdorferi,* would decrease the number of blacklegged ticks that are able to transmit Lyme disease to humans (Ginsberg et al, 2021). Future research could investigate host-tick relationships and the effect of location on these relationships, thereby affecting the spread of Lyme disease. Another possible explanation for the shift in trends from increasing to decreasing could be increases in local policies and interventions to reduce Lyme disease cases, such as education policies and DEET insect repellent (CDC, 2022).

A limitation to this study was the lack of monthly or weekly data at the state/locality level. This data would allow for a more in-depth fitted model per location, as well as for the total US, and would possibly increase the accuracy of the total Lyme disease case predictions per state/locality. Another limitation was the use of automated model fitting; however, it worked fine in this example as the ACF, PACF and QQ-plot did not indicate a lack-of-fit for the logarithmic cases model. There were also some limitations to the Lyme disease case data used; including under-reporting and misclassification, which is why true Lyme disease case numbers are estimated at much higher than those reported (CDC data, 2022). Also, this data classifies a case by the patient's state/location of residence and not exposure (CDC data, 2022). Lastly, the data is collected by state/locality local departments, and this depends on the state's ability to classify cases, which can vary even year to year due to budget or personnel changes (CDC data, 2022). This could lead to a change in reported cases that is not representative of an actual change in Lyme disease incidence (CDC data, 2022).

In conclusion, the 2022 to 2024 US total Lyme disease forecast shows no increasing or decreasing trend in Lyme disease cases; however, this could change with temperatures increasing due to climate change. At the state/locality level there is a more recent decreasing trend in Lyme disease cases, possibly indicative of the effect of education policies and interventions in these states. However, future research should investigate each location trend with temperature and possible tick-host interactions to get a more in-depth understanding of the change in these trends.

References

Achim Zeileis, Torsten Hothorn (2002). Diagnostic Checking in Regression Relationships. R News 2(3), 7-10. URL https://CRAN.R-project.org/doc/Rnews/

Beard, C.B., R.J. Eisen, C.M. Barker, J.F. Garofalo, M. Hahn, M. Hayden, A.J. Monaghan, N.H. Ogden, and P.J. Schramm. 2016. Chapter 5: Vector-borne diseases. In: The impacts of climate change on human health in the United States: A scientific assessment. U.S. Global Change Research Program. https://health2016.globalchange.gov

Box, G.E.P., Jenkins, G.M. (1970) Time Series Analysis*, Forecasting and Control.* Sen Francisco, Holden-Day

CDC Center for Disease Control and Prevention, Lyme Disease, published January 2022, retrieved on November 25, 2023 from https://www.cdc.gov/lyme/index.html

CDC Center for Disease Control and Prevention, Lyme Disease Surveillance Data, published October 2022, retrieved on November 11, 2023 from https://www.cdc.gov/lyme/datasurveillance/surveillance-data.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Flyme%2Fdatasurveillance%2Frecent-surveillance-data.html

Chatfield, C. (1996) *The Analysis of Time Series* (5th edn).

EPA United States Environmental Protection Agency, Climate Change Indicators: Lyme Disease, published November 2023, retrieved on November 25, 2023 from https://www.epa.gov/climate-indicators/climate-change-indicators-lyme-disease

Ginsberg, H. S., Hickling, G. J., Burke, R. L., Ogden, N. H., Beati, L., LeBrun, R. A., Arsnoe, I. M., Gerhold, R., Han, S., Jackson, K., Maestas, L., Moody, T., Pang, G., Ross, B., Rulison, E. L., & Tsao, J. I. (2021). Why Lyme disease is common in the northern US, but rare in the south: The roles of host choice, host-seeking behavior, and tick density. *PLoS biology, 19*(1), e3001066. https://doi.org/10.1371/journal.pbio.3001066

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2023). _forecast: Forecasting functions for time series and linear models_. R package version 8.21.1, <https://pkg.robjhyndman.com/forecast/>.

Komaroff, A. L., & Lipkin, W. I. (2023). ME/CFS and Long COVID share similar symptoms and biological abnormalities: road map to the literature. *Frontiers in medicine*, *10*, 1187163. https://doi.org/10.3389/fmed.2023.1187163

Ljung, G.M., Box, G.E.P. (1978) On a Measure of a Lack of Fit in Time Series Models. *Biometrika* 65(2):297–303.

NOAA National Centers for Environmental information, Climate at a Glance: National Time Series, published November 2023, retrieved on November 25, 2023 from https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/national/time-series

Rebman, A. W., & Aucott, J. N. (2020). Post-treatment Lyme Disease as a Model for Persistent Symptoms in Lyme Disease. *Frontiers in medicine*, *7*, 57. https://doi.org/10.3389/fmed.2020.00057

Skar, G. L., & Simonsen, K. A. Lyme Disease (Updated 2023). StatPearls [Internet]. *Treasure Island (FL): StatPearls Publishing*. Available from: https://www.ncbi.nlm.nih.gov/books/NBK431066/

Figures

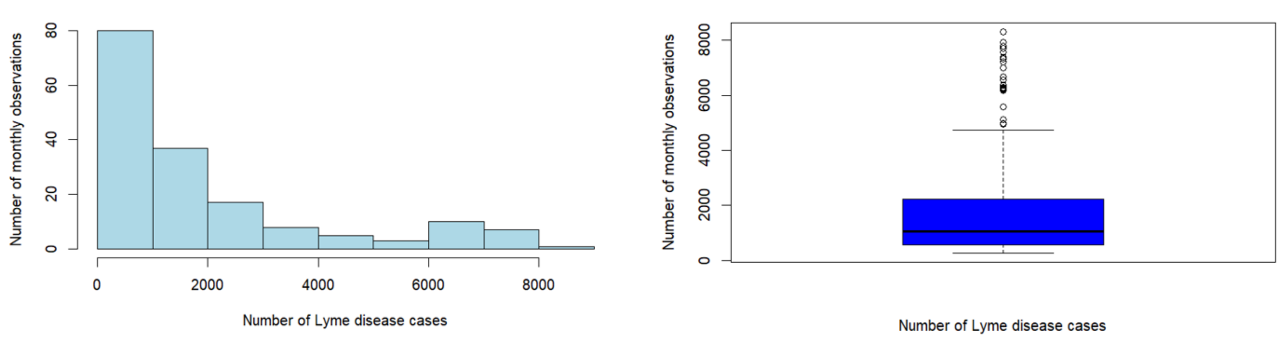

Figure 1: Histogram (left) and boxplot (right) of US total monthly Lyme disease case data from 2008 to 2021.

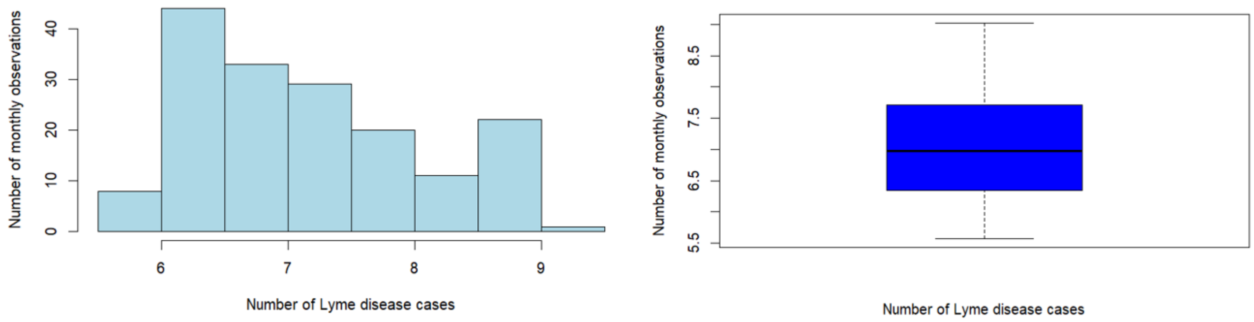

Figure 2: Histogram (left) and boxplot (right) of logarithmic US total monthly Lyme disease case data from 2008 to 2021.

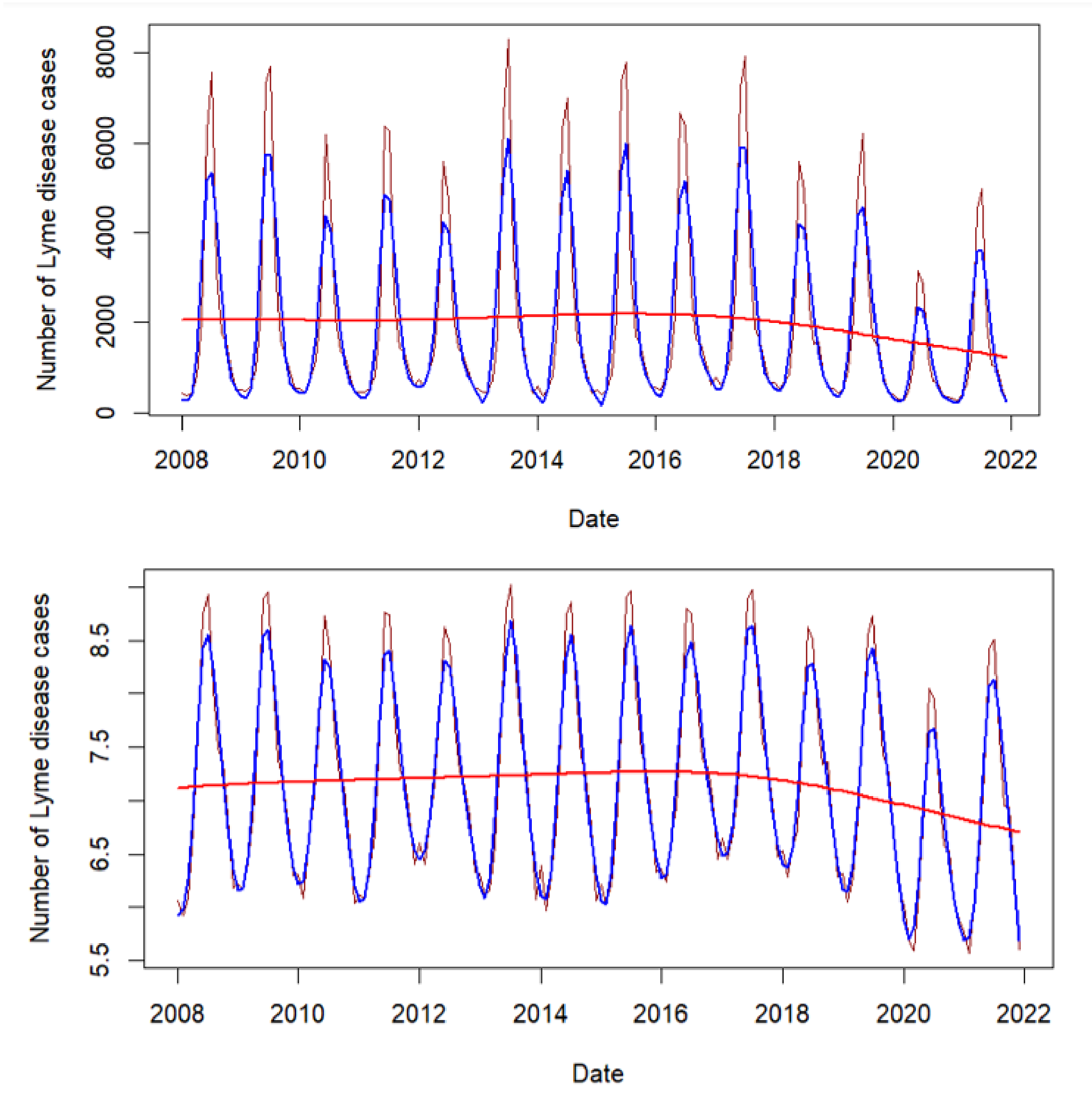

Figure 3: US Lyme disease cases with smooth splines visualizing possible trend and seasonal components (top), and logarithmic US Lyme disease cases with smooth splines visualizing possible trend and seasonal components (bottom).

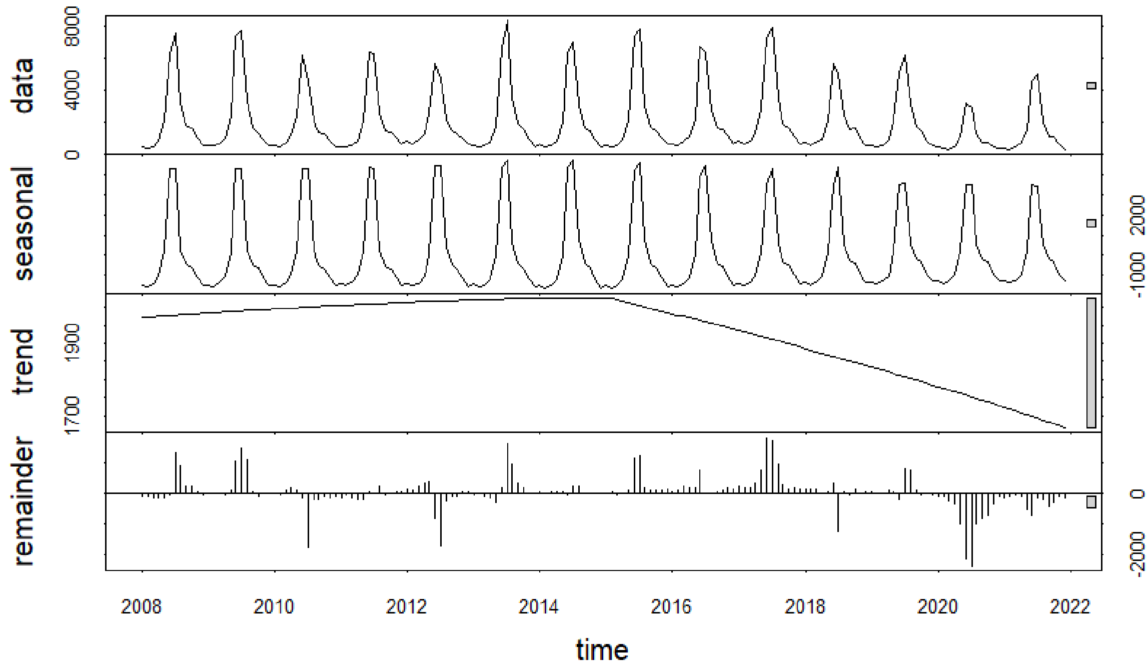

Figure 4: STL analysis plots for US monthly Lyme disease case data from 2008 to 2021.

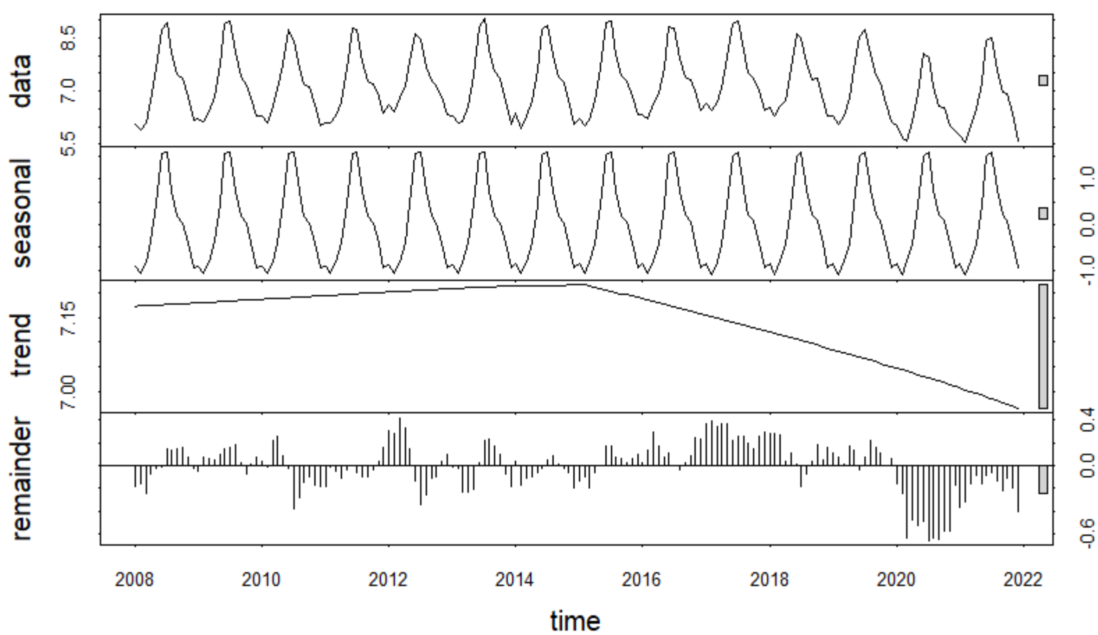

Figure 5: STL analysis plots for logarithmic US monthly Lyme disease case data from 2008 to 2021.

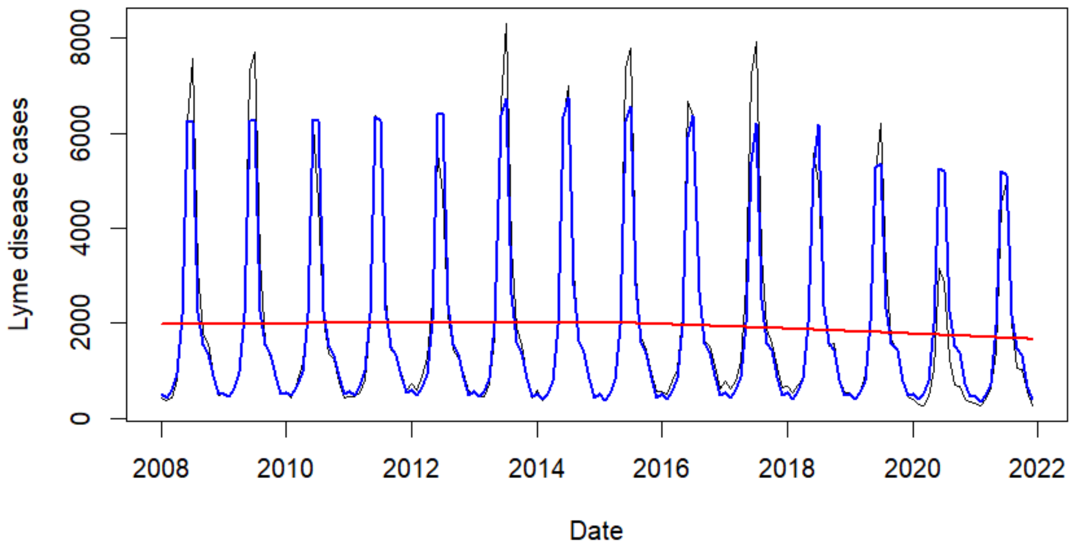

Figure 6: US monthly Lyme disease case data plotted with trend and seasonal components of the corresponding STL analysis.

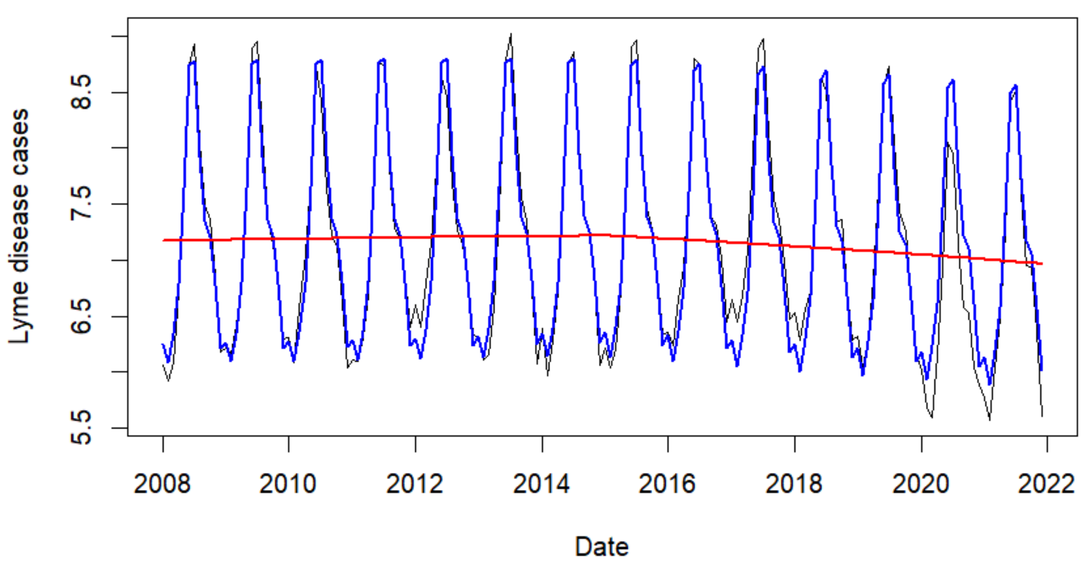

Figure 7: Logarithmic US monthly Lyme disease case data plotted with trend and seasonal components of the corresponding STL analysis.

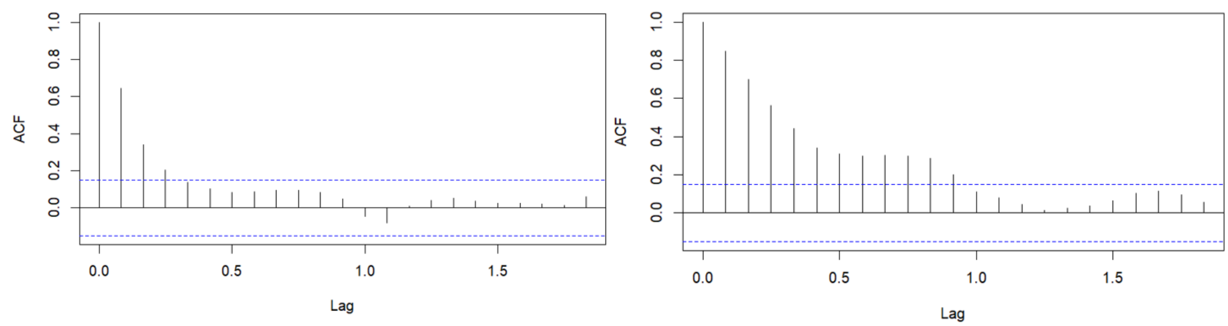

Figure 8: ACF plots for STL analysis of US monthly Lyme disease case data (left) and of the logarithmic US monthly Lyme disease case data (right).

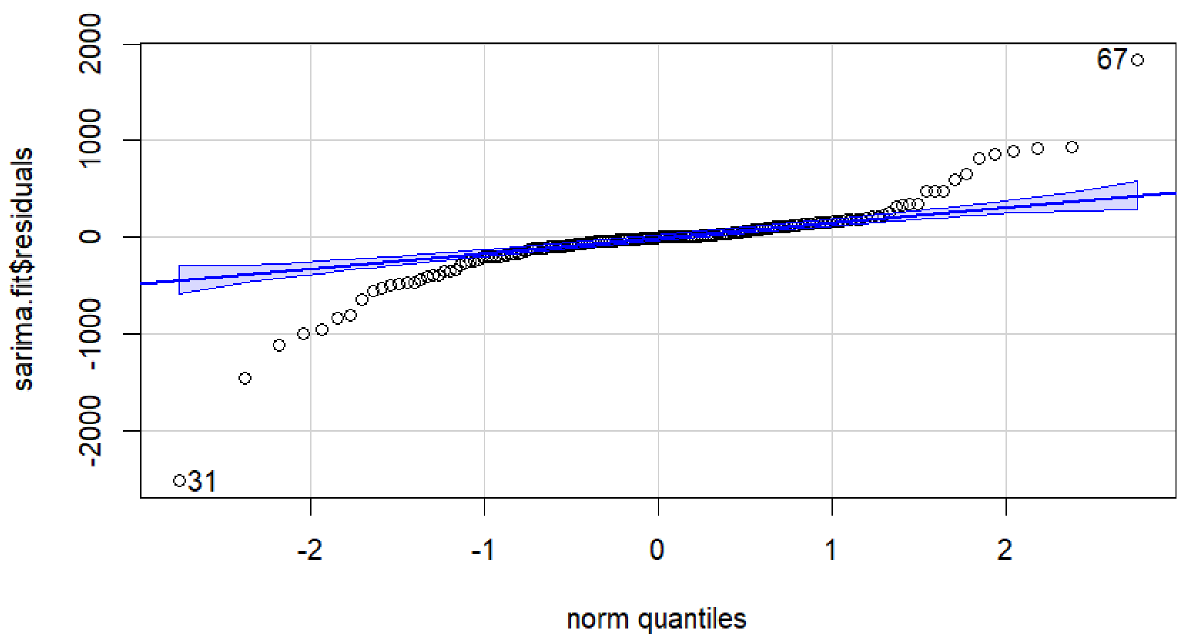

Figure 9: QQ plot for the SARIMA model fit to US monthly Lyme disease case data from 2008 to 2021.

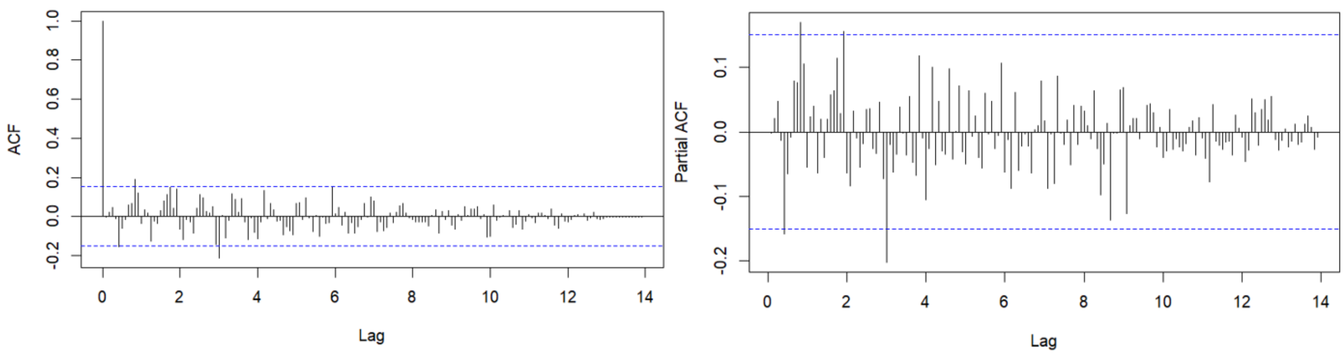

Figure 10: ACF (left) and PACF (right) plots for the SARIMA model fit to logarithmic US monthly Lyme disease case data from 2008 to 2021.

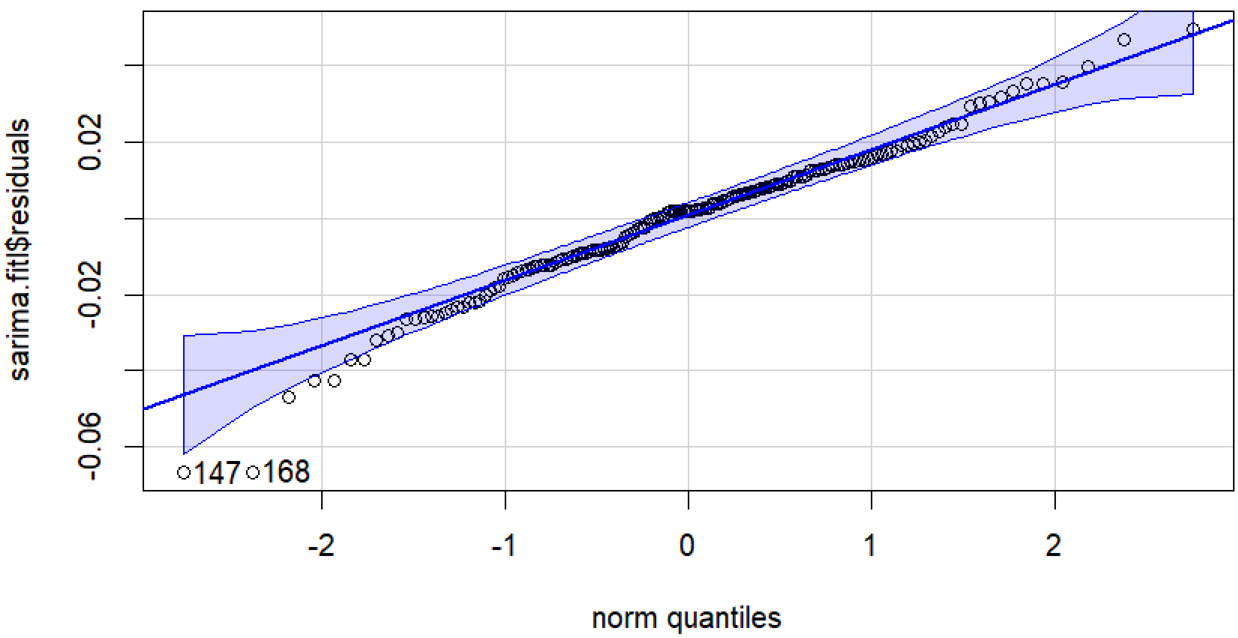

Figure 11: QQ plot for the SARIMA model fit to logarithmic US monthly Lyme disease case data from 2008 to 2021.

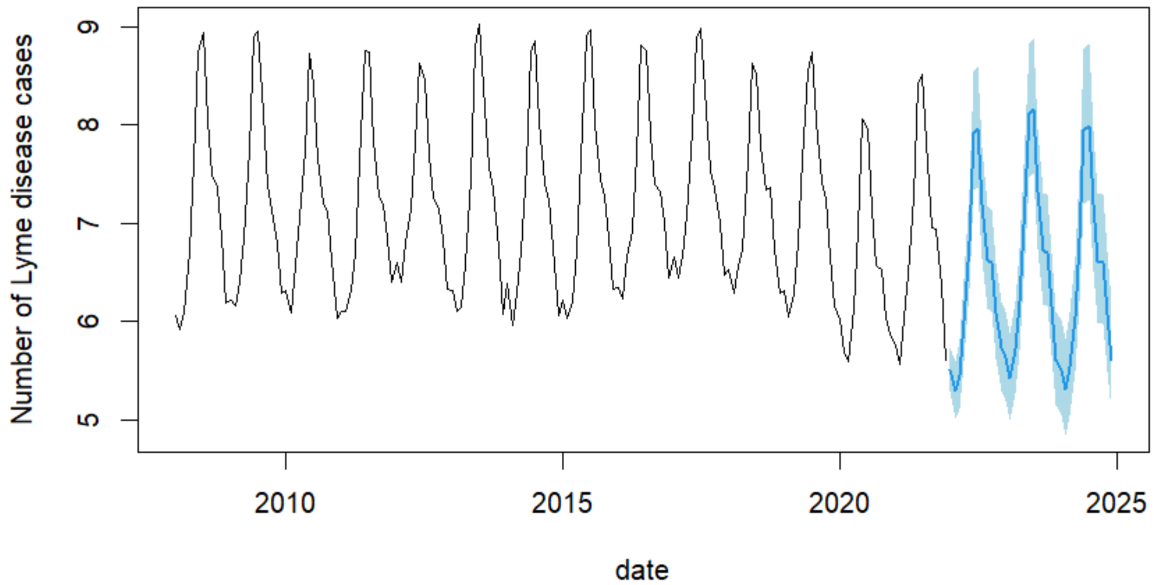

Figure 12: Logarithmic US monthly Lyme disease case data with projected values for 2022 to 2024.

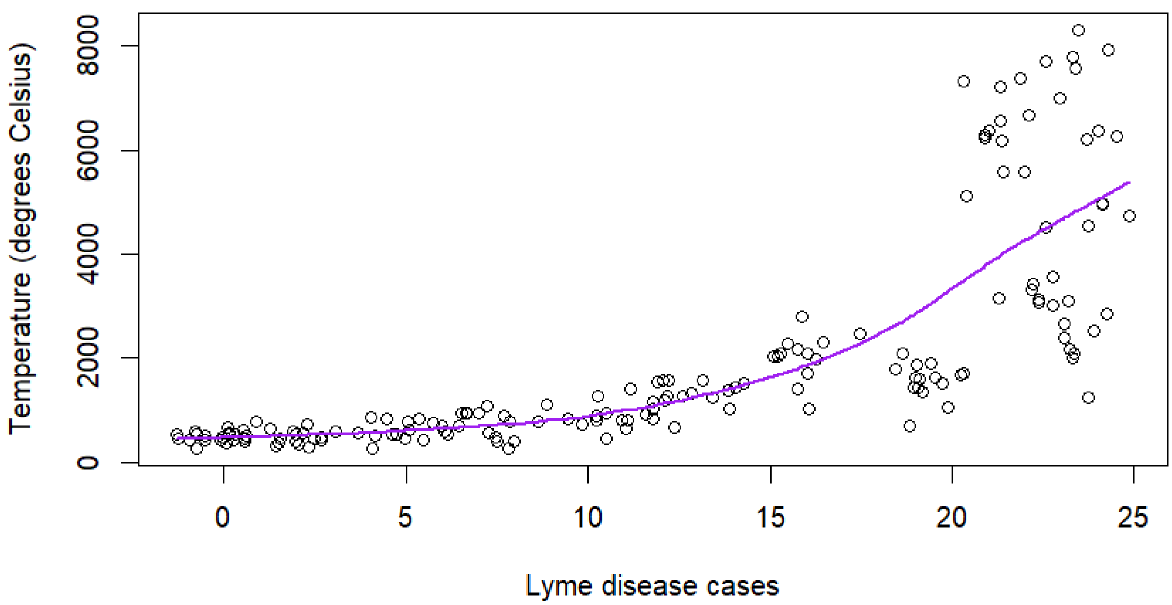

Figure 13: Scatter plot of US monthly Lyme disease cases versus US monthly temperature (degrees Celsius), with a smooth spline visualizing possible correlation.

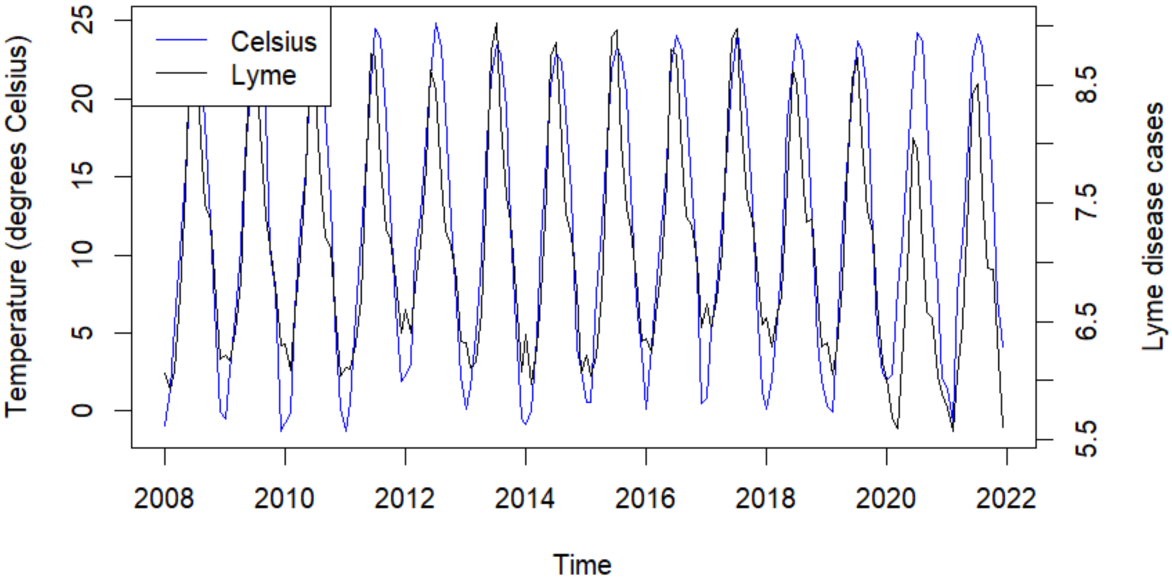

Figure 14: Logarithmic US monthly Lyme disease case data plotted with US monthly temperature data (degrees Celsius).

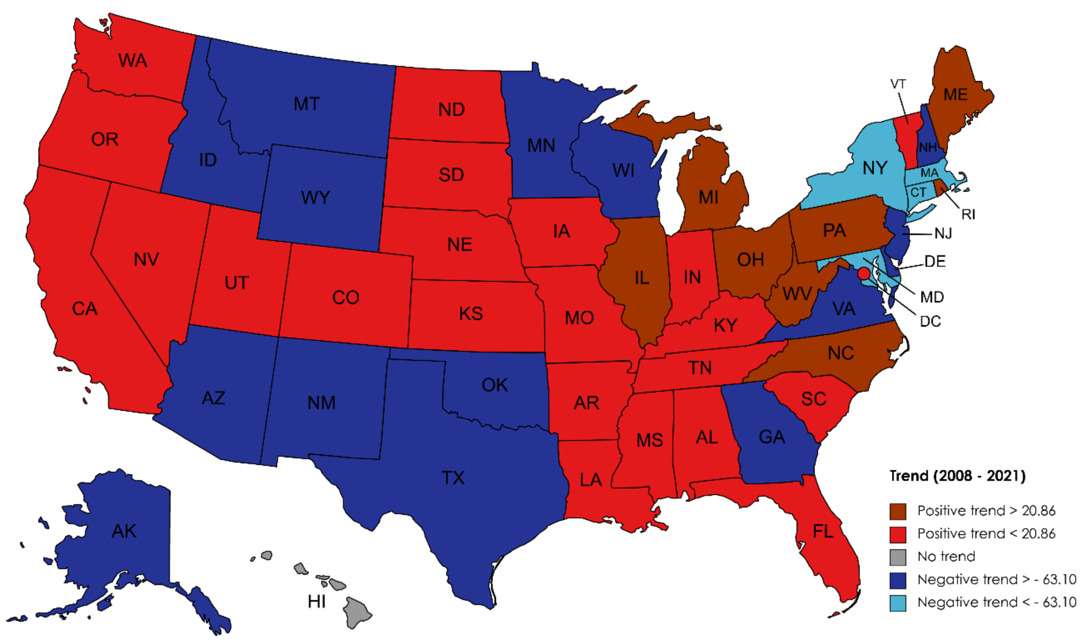

Figure 15: Map of US states and localities with positive (red), negative (blue), or no (grey) trends from the linear regression of yearly Lyme disease case data per state/locality from 2008 to 2021.

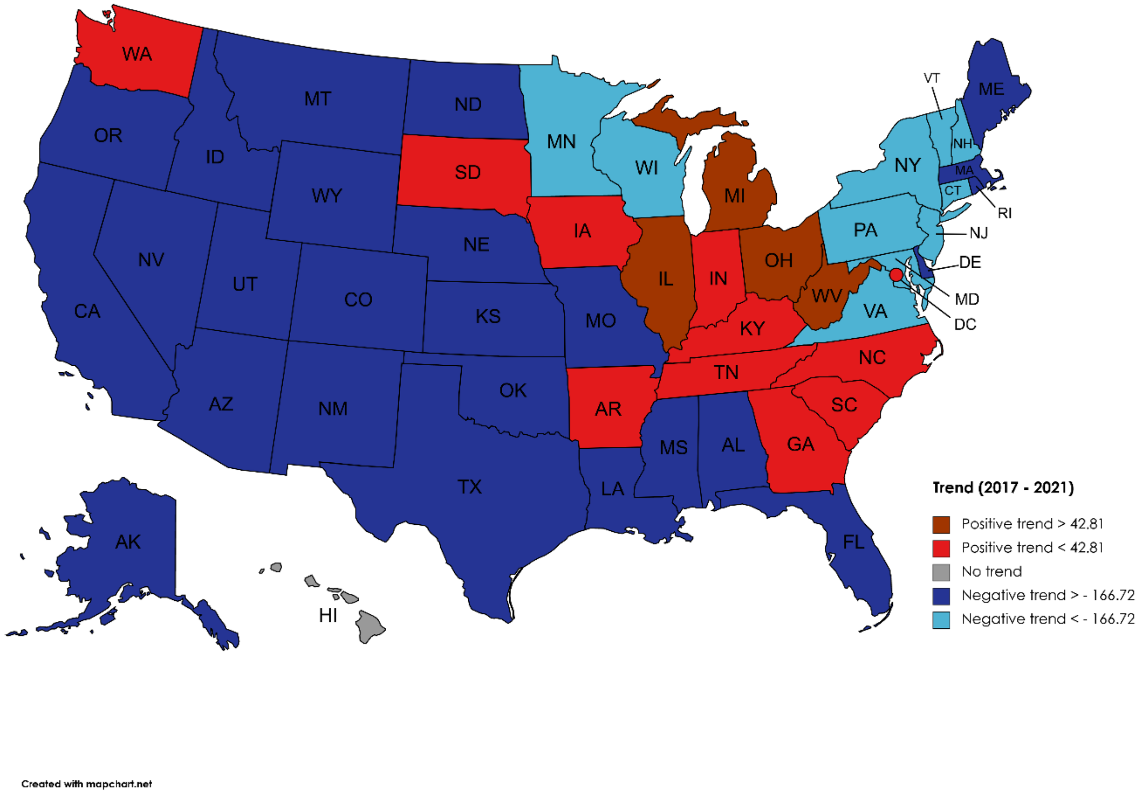

Figure 16: Map of US states and localities with positive (red), negative (blue), or no (grey) trends from the linear regression of yearly Lyme disease case data per state/locality from 2017 to 2021.

Tables

Table 1: Projected Lyme disease total cases for each state/locality in the US for 2024, with a 95 percent confidence interval, based on the 13-year (2008 to 2021) linear regression model.

State/locality	Point Forecast	95% CI (lower)	95% CI (upper)
Alabama	59.58242	16.06266	103.1022
Alaska	4.024176	0	15.64466
Arizona	10.37363	0	35.50324
Arkansas	14.24615	4.352588	24.13972
California	111.9319	33.22494	190.6388
Colorado	2.323077	0	8.789905
Connecticut	0	0	764.1465
Delaware	200.9033	0	583.8312
District of Columbia	123.1868	36.85506	209.5186
Florida	214.367	132.1051	296.629
Georgia	15.48352	0	51.62284
Hawaii	0	0	0

Idaho	10.85714	0	25.51076
Illinois	477.3978	307.968	646.8276
Indiana	278.7582	199.0246	358.4919
Iowa	396.1912	283.9077	508.4747
Kansas	34.85055	12.0448	57.6563
Kentucky	46.7978	11.23381	82.3618
Louisiana	7.015385	0	15.47027
Maine	1909.057	1160.505	2657.609
Maryland	950.2352	109.8497	1790.621
Massachusetts	0	0	2316.988
Michigan	646.156	308.8233	983.4888
Minnesota	1397.495	0	2983.456
Mississippi	2.725275	0	7.30624
Missouri	12.94945	3.533839	22.36506
Montana	8.96044	0	23.55654
Nebraska	11.14725	1.924776	20.36973
Nevada	15.18462	1.376484	28.99275
New Hampshire	597.8945	0	1753.936
New Jersey	3560.833	1666.941	5454.725
New Mexico	0.195604	0	6.96015
New York	2365.163	0	5355.423
North Carolina	403.1231	310.1342	496.1119
North Dakota	46.94725	21.6196	72.2749
Ohio	570.7978	374.2989	767.2967
Oklahoma	0	0	2.240627
Oregon	75.08571	40.09641	110.075
Pennsylvania	8764.607	697.5513	16831.66
Rhode Island	1418.473	875.5126	1961.432
South Carolina	57.79121	28.40024	87.18217
South Dakota	14.88791	7.973131	21.80269
Tennessee	44.05714	17.87625	70.23803
Texas	0	0	83.055
Utah	25.95604	11.90725	40.00484
Vermont	605.5582	0	1456.46
Virginia	1073.178	243.4804	1902.876
Washington	39.15165	18.06232	60.24097
West Virginia	1420.952	690.3793	2151.524
Wisconsin	1700.646	-89.8115	3491.104
Wyoming	1.734066	0	6.163159

Table 2: Projected Lyme disease total cases for each state/locality in the US for 2024, with a 95 percent confidence interval, based on the 5-year (2017 to 2021) linear regression model.

State/locality	Point Forecast	95% CI (lower)	95% CI (upper)
Alabama	40.6	0	176.4195
Alaska	0	0	4.583475
Arizona	0	0	24.8377
Arkansas	19.2	0	52.42978
California	45.9	0	269.905
Colorado	0	0	17.46166
Connecticut	0	0	247.3322
Delaware	0	0	694.2531
District of Columbia	122.3	26.13762	218.4624
Florida	130	0	351.2134
Georgia	55.9	33.46357	78.33644
Hawaii	0	0	0
Idaho	0	0	19.96071
Illinois	632.3	110.3988	1154.201
Indiana	402.9	224.7392	581.0608
Iowa	382.4	159.6897	605.1103
Kansas	4.3	0	41.48563
Kentucky	53.4	28.71056	78.08944
Louisiana	0	0	16.86758
Maine	1133.5	0	3779.715
Maryland	44.6	0	1197.249
Massachusetts	0	0	715.798
Michigan	1133.2	343.4764	1922.924
Minnesota	271.3	0	5871.969
Mississippi	0.7	0	11.82036
Missouri	3.8	0	28.59042
Montana	7.6	0	36.05043
Nebraska	1.8	0	21.28669
Nevada	8.8	0	46.99643
New Hampshire	0	0	2491.499
New Jersey	1468	0	5397.938
New Mexico	0.8	0	18.54196
New York	808.4	0	5517.104
North Carolina	357.6	20.25088	694.9491
North Dakota	18.9	0	67.75755
Ohio	783.8	401.5546	1166.045
Oklahoma	0	0	1.435277
Oregon	40.9	0	145.3319
Pennsylvania	0	0	3376.786
Rhode Island	718.6	176.671	1260.529
South Carolina	105.7	60.61884	150.7812
South Dakota	15.1	0	38.30278

Tennessee	57.7	0	124.0749
Texas	0	0	71.26452
Utah	7.1	0	29.53644
Vermont	0	0	1505.638
Virginia	0	0	1146.31
Washington	38.6	0	120.638
West Virginia	2353.2	953.6846	3752.715
Wisconsin	1119.4	0	4440.459
Wyoming	0.2	0	15.68645