Predicting Future Earthquake Count Using AutoRegressive Integrated Moving Average Model

Xi Zheng(1005153628)

Abstract

Earthquakes as a highly destructive natural disaster, have taken tens of thousands of lives. High-magnitude earthquakes occur all throughout the planet, and each time they occur, they also create a chain that induces other disasters. In this report, four autoregressive integrated moving average models are proposed by using data of earthquake(magnitude 7 and above) counts from 1900 to 2006 with the goal of predicting the frequency of earthquakes of magnitude 7 and above in future years. The final model AR(1) forecast 19 earthquakes with magnitude 7 and above for 2021, and the actual number[1] of earthquakes with magnitude 7 and above for 2021 till December 17 is 18. The model predictions for other years are not very accurate, but the model shows that the number of large earthquakes will have a decreasing trend in this century.

Key Words

AutoRegressive Integrated Moving Average(ARIMA) Akaike information criterion(AIC)

Bayesian Information Criterion(BIC) Spectral Analysis

Introduction

Earthquakes often last only a few dozen seconds to a few minutes, but that is enough time to kill thousands of people and destroy countless buildings. High-magnitude earthquakes occur all throughout the planet, and each time they occur, they also create a chain that induces other disasters. For nearly 100 years, there have been about 19 big earthquakes each year. The Haiti Earthquake(magnitude 7) happened in 2010 killed 316,000 people and the Great Tangshan Earthquake(magnitude 7.5) killed about 242,769 people[2]. Predicting earthquakes has been the goal of mankind since ancient times. As early as 132 A.D.[3], Zhang

Heng, a scientist in the Eastern Han Dynasty of China, made the world's first "seismograph". Today, about two thousand years later, people is still making efforts to predict earthquakes.

The EQcount data in astsa library records annual counts of major earthquakes (magnitude 7 and above) in the world between 1900 and 2006. In this report, we use the EQcount build four ARIMA models and aim to predict the number of earthquakes magnitude 7 and above in future years. Specific statitical methods, results, and discussion of the results are in the following sections.

Statistical Methods

There are totally 107 data points in the EQcount dataset, figure 1 shows the number of earthquake with magnitude 7 and above between 1900 to 2006. The orange line in figure1 represents the mean count which is equal to 19.36. The maximum count in EQcount is 41 and the minimum count is 6. We can clearly see that there is no seasonal or other obvious trend in the plot. EQcount is approximatly stationary since it has a roughly constant mean and variance.

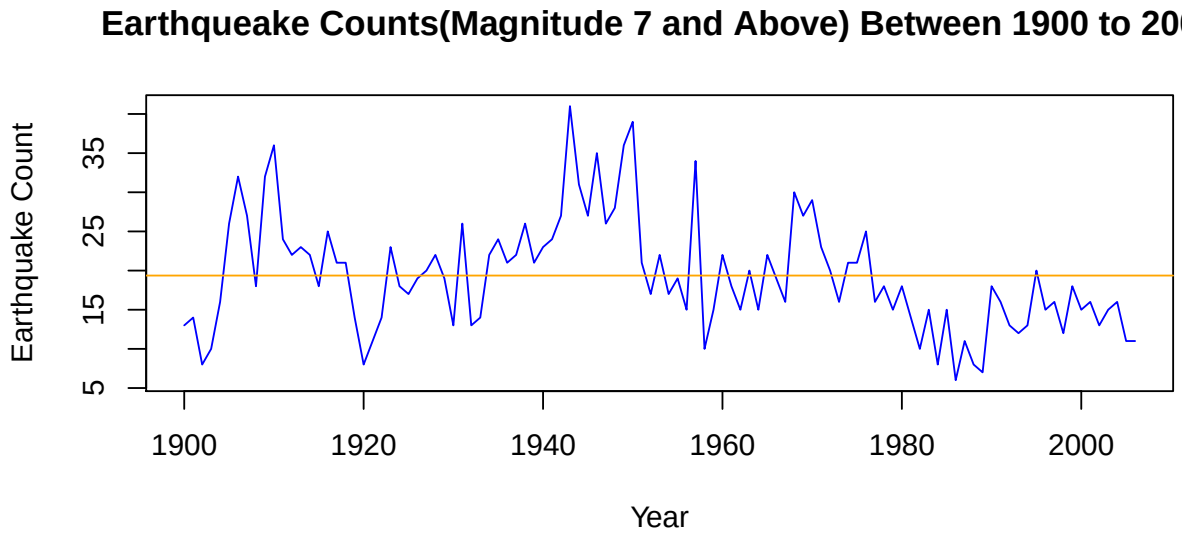


Figure 1. Number of earthqueakes magnitude 7 and above between 1900 to 2006

However, in order to get the best model, we decide to difference the EQcount data once. The detrended earthquake counts plot in figure2 looks very stationary, it has a mean of zero and constant variance.

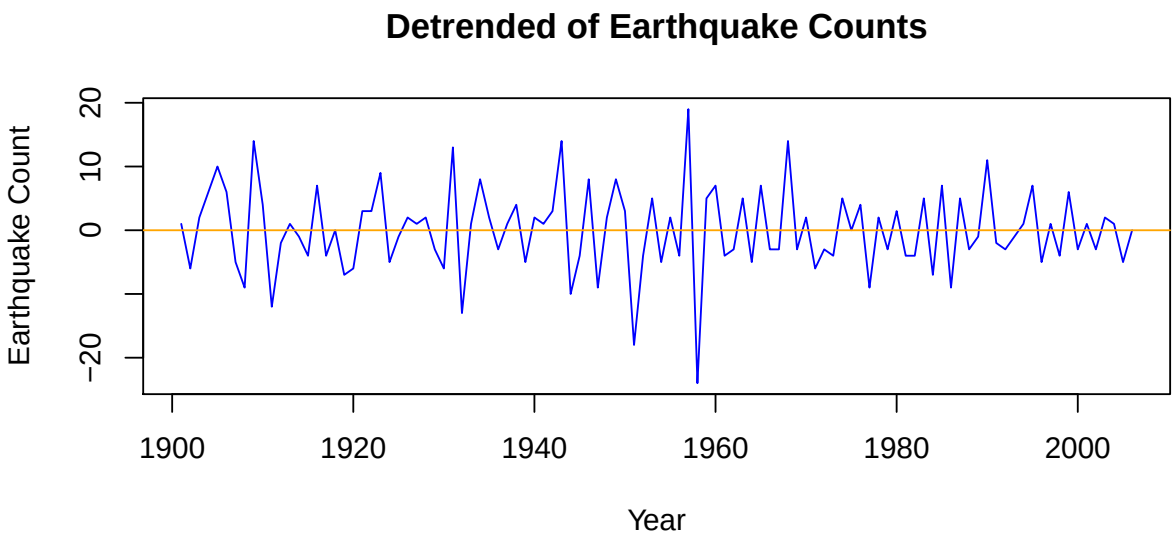


Figure 2. Detrended earthquake counts magnitude 7 and above between 1900 to 2006

We could see that PACF in figure3 is cutting off at lag 1, and ACF is cutting off at lag 7. Therefore we firstly propose two models, the first one is AR(1), the second model is MA(7). After doing the differencing, we could see that PACF in figure4 is cutting off at lag 2 and ACF is cutting at lag 1. Thus we propose our third model ARIMA (2, 1, 0) and the forth model ARIMA = (0, 1, 1).

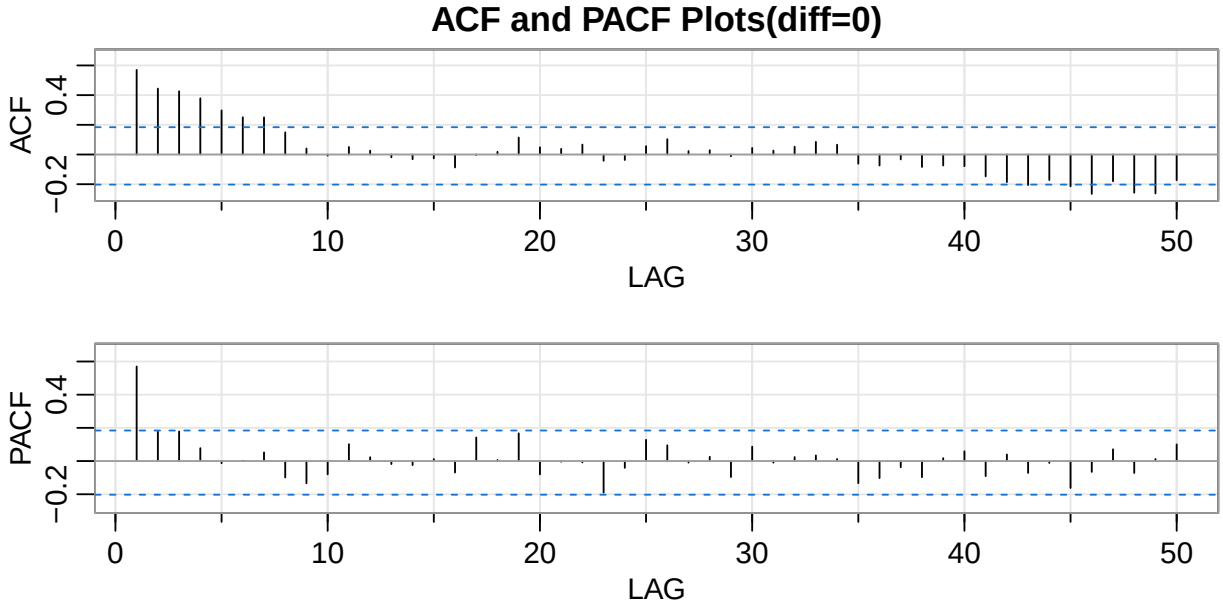


Figure 3. ACF and PACF plots for EQcount data

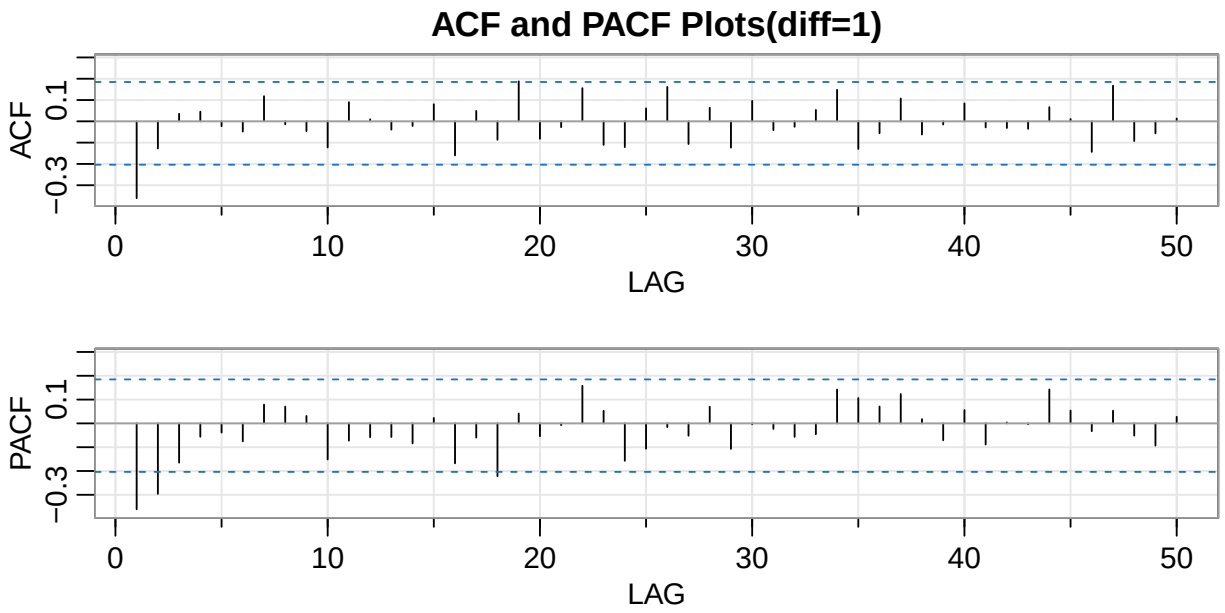


Figure 4. ACF and PACF plots for EQcount data after one differencing

Results

Table 1-4 are the output results of our 4 proposed models. We can clearly see that all parameters in AR(1) model are significant, and parameters expect constant terms in ARIMA(2,1,0), ARIMA(0,1,0) are all significant, their p values are all smaller than the significant level of 0.05. However, few paramters in MA(7) are not significant, so we drop the MA(7) model and we are now having three models left.

Table 1. Parameter estimate and AIC values for proposed AR(1)

AR(1)	Estimate	p-value		
ar1	0.5764	0	AIC	6.428499
xmean	19.1819	0	AICc	6.429577
			BIC	6.503438

Table 2. Parameter estimate and AIC values for proposed MA(7)

MA(7)	Estimate	p-value		
ma1	0.3965	0.0001	AIC	6.448303
ma2	0.2257	0.0282	AICc	6.462036
ma3	0.3186	0.0039	BIC	6.67312
ma4	0.2831	0.0081		
ma5	0.2112	0.0838		
ma6	0.1753	0.0694		
ma7	0.2307	0.0235		
xmean	18.9692	0.0000		

Table 3. Parameter estimate and AIC values for proposed ARIMA(2,1,0)

ARIMA(2,1,0)	Estimate	p-value		
ar1	-0.4644	0.0000	AIC	6.45261
ar2	-0.2944	0.0019	AICc	6.45483
constant	-0.0056	0.9862	BIC	6.553118

Table 4. Parameter estimate and AIC values for proposed ARIMA(0,1,1)

ARIMA(0,1,1)	Estimate	p-value		
ma1	-0.5762	0.0000	AIC	6.415915
constant	-0.0090	0.9703	AICc	6.417014
			BIC	6.491295

Figure 5 to 7 are the diagnostic plots for AR(1), ARIMA(2,1,0) and ARIMA(0,1,1) models. For both there models, the standardized residuals all behave as sequences with mean zero and variance one, no any obvious departures are found. For both three models, the normal Q-Q plot shows normality. For both three models, the ACF plots looks close to white noise. For both three models, the p values for Ljung-Box statistic are all above the baseline, so we can not reject the null hypothesis that residuals are uncorrealted. Therefore, they all pass the diagnostic checks.

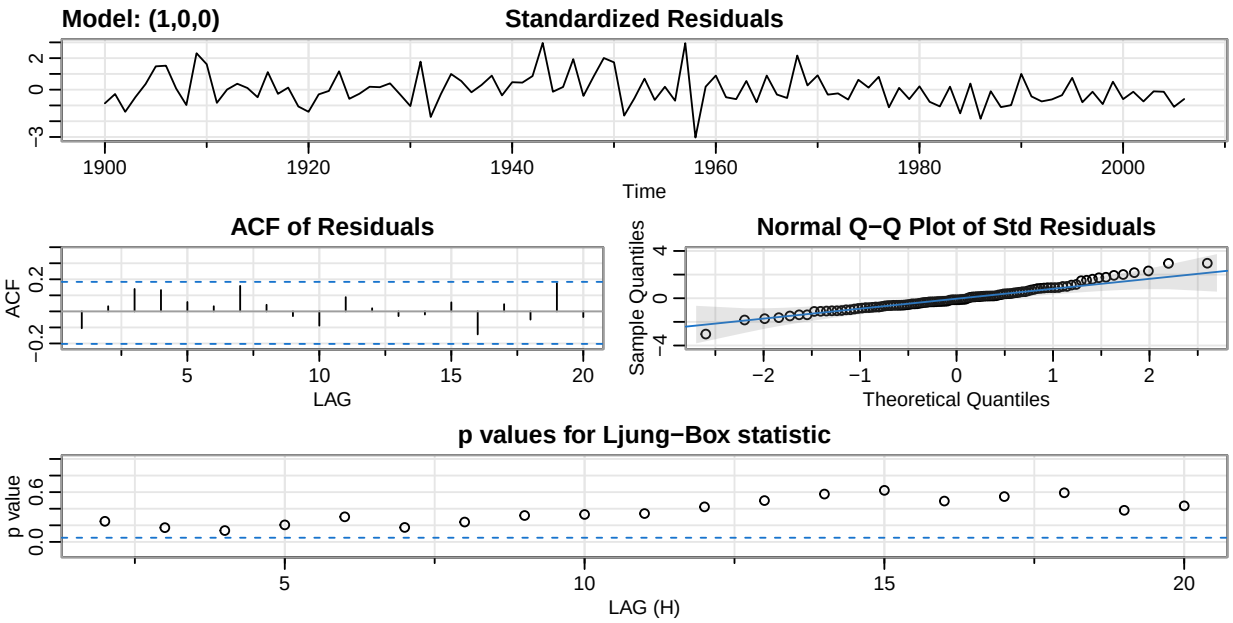


Figure 5. Diagnostic plots for AR(1)

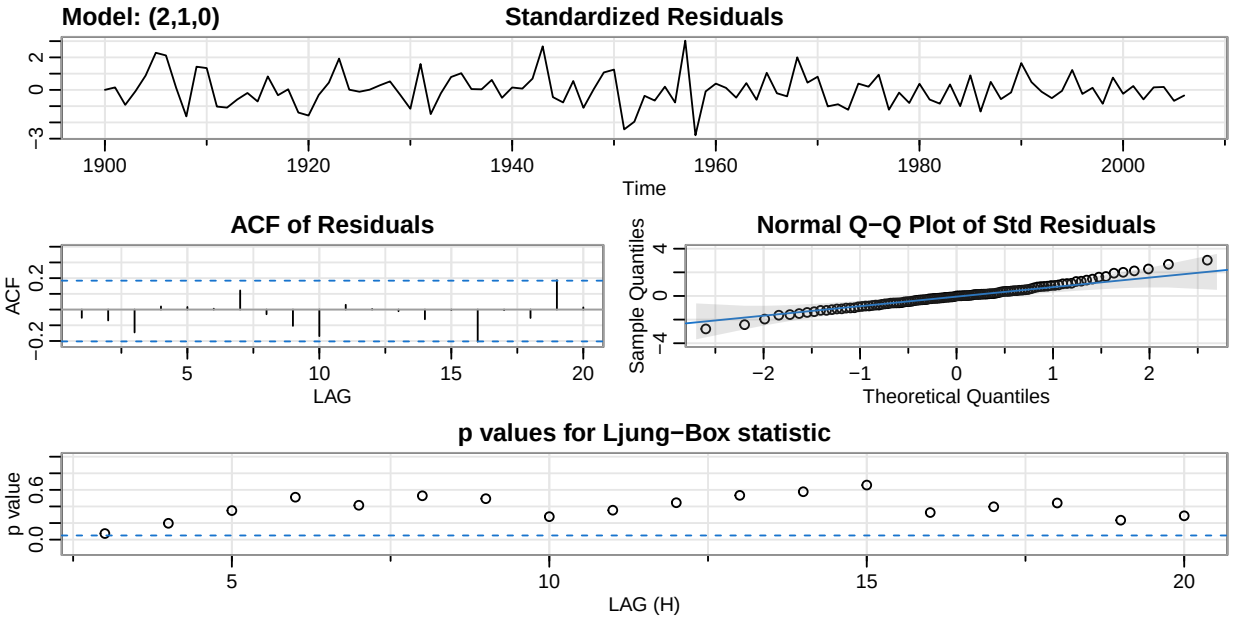


Figure 6. Diagnostic plots for ARIMA(2,1,0)

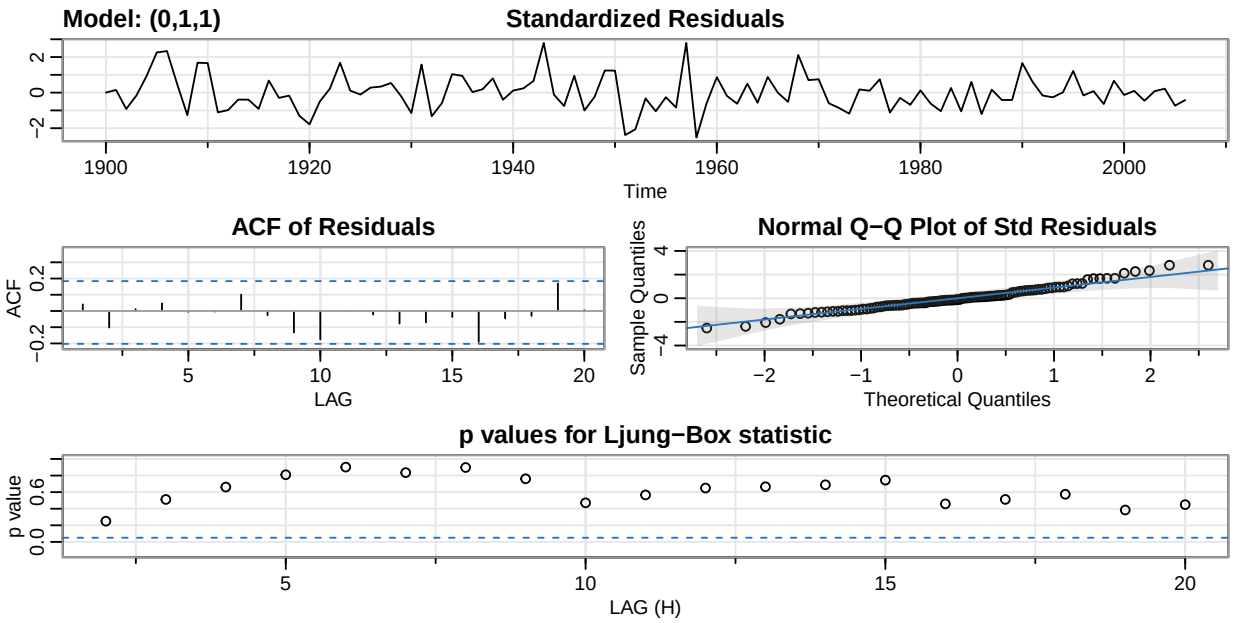


Figure 7. Diagnostic plots for ARIMA(0,1,1)\newline

Even though ARIMA(0,1,1) and ARIMA(2,1,0) all pass the diagnotic checks, their constant term is not significant. Thus, we propose AR(1) modelas our final model because all its parameters are significant, and it has the second smallest AIC, AICc, BIC values(values can be founde in table 1 to 4 above). The final equation is:

$$\hat{x}_t = 19.1819 * (1 - 0.5764) + 0.5764\hat{x}_{t-1} + \hat{w}_t$$

Using the above model to forecast for future 15 years, we predict that there will be 19 major earthquakes in the year of 2021. Figure8 shows the predicted plot.

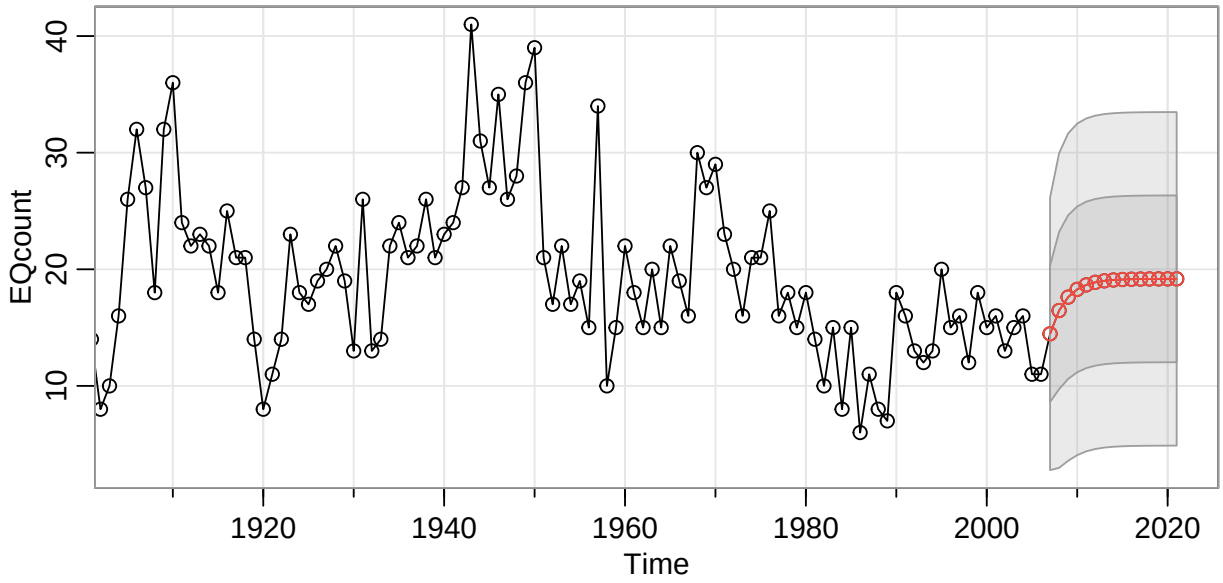
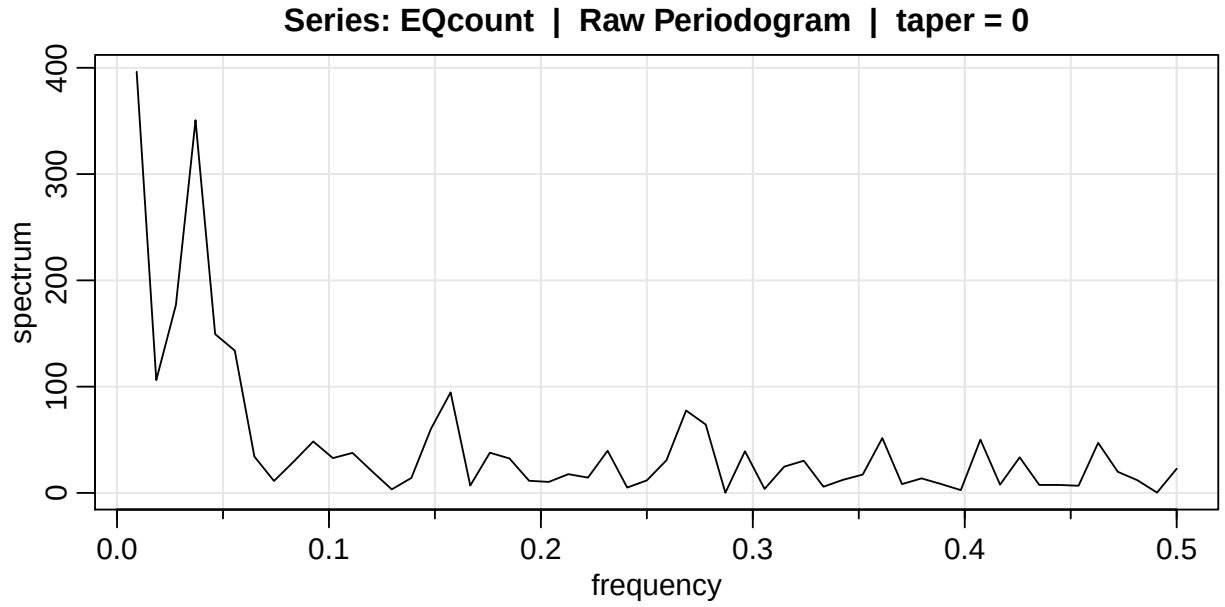


Figure 8. Predicted plot for AR(1) model for future 15 years

The predict values and 95% confidance intervals as follows:

```
##    Predicts    Upper    Lower
## 1  14.46614 24.07838 4.853886
## 2  16.46389 27.55841 5.369360
## 3  17.61532 29.16021 6.070421
## 4  18.27896 29.96963 6.588291
## 5  18.66146 30.40015 6.922766
## 6  18.88192 30.63652 7.127314
## 7  19.00898 30.76886 7.249098
## 8  19.08222 30.84385 7.320579
## 9  19.12443 30.88664 7.362206
## 10 19.14875 30.91117 7.386341
## 11 19.16278 30.92525 7.400299
## 12 19.17086 30.93336 7.408359
## 13 19.17552 30.93802 7.413010
## 14 19.17820 30.94071 7.415692
## 15 19.17975 30.94226 7.417239
```

We also perform a spectral analysis and identify the first three predominant periods and get the confidence intervals as outputs shown below. The confidence intervals are extremely wide, so we are unable to establish significance of the peak.




```
##      frequency period spectrum
## [1,]    0.0093    108 396.2778
## [2,]    0.0370     27 350.7439
## [3,]    0.0278     36 176.8623
```

```
##            CIl      CIu
## [1,] 132.28078 7725.723
## [2,] 117.08119 6838.007
## [3,]  59.03809 3448.059
```

Discussion

Our final model AR(1) forecast 19 earthquakes with magnitude 7 and above for 2021, and the actual number of earthquakes with magnitude 7 and above for 2021 till December 17 is 18. The mean of our predicted values for 15 years is 18 and the actual mean count from 2007 to 2021 is 15. Compared with the mean of 19 in 20 century, we predicted a downward trend in the number of large earthquakes. This suggests that the number of big earthquakes will be less frequent in the coming period than in the last century.

Despite the fact that the predicted value for 2021 is quite close to the actual number, the shape of our predicted curve, as shown in figure 9, differs from the actual curve, indicating that our model is inadequate. Our model does have some drawbacks. To begin with, the size of our training data is inadequate. Technology was not as advanced as it is today prior to the twentieth century. As a result, the number of earthquakes that occurred prior to the twentieth century is unknown. Our training set was too small because we only had 106 data points. Another disadvantage, in addition to the training data size, is that our ARIMA model does not account for seasonal behaviour. However, since the earthquake may be linked to seasonal fluctuations, we should try to make some improvements to create a seasonal ARIMA model or apply a sine/cosine transformation.

As a result, future researchers should pay more attention to capturing the pattern of this dataset, as the AR(1) model is insufficiently accurate for future prediction. Another recommendation is to reduce the targeted magnitude because smaller earthquakes occur more frequently than huge earthquakes, hence smaller earthquakes should have a higher training size.

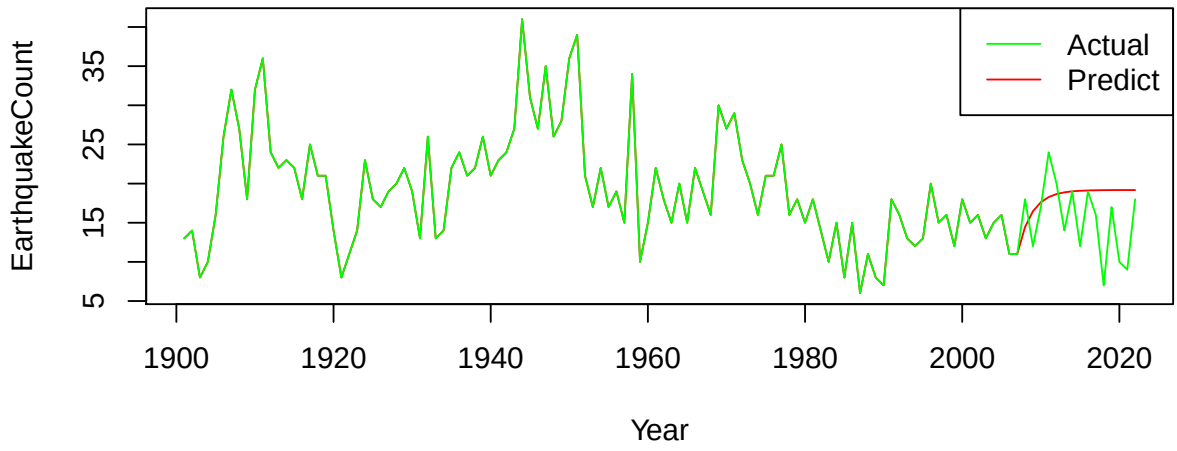


Figure 9. Predicted plot for AR(1) model for future 15 years vs acutal data

Reference

[1] Lists, maps, and Statistics. Lists, Maps, and Statistics | U.S. Geological Survey. (n.d.). Retrieved December 17, 2021, from https://www.usgs.gov/programs/earthquake-hazards/lists-maps-and-statistics

[2] Encyclopædia Britannica, inc. (n.d.). The 6 deadliest earthquakes since 1950. Encyclopædia Britannica. Retrieved December 17, 2021, from https://www.britannica.com/list/6-deadliest-earthquakes

[3] How are earthquakes studied?: UPSeis. Michigan Technological University. (n.d.). Retrieved December 17, 2021, from https://www.mtu.edu/geo/community/seismology/learn/earthquake-study/