Adam Chaffee
March 21, 2017

A Time Series Analysis of Flight Delays at LAX

Introduction

With the reduction in fuel price and increased airfare accessibility to the middle class, the population of airline passengers is growing at a record pace. As this population continues to grow, airports are becoming an increasingly important component of the global economy. Last year, U.S. airlines carried a record 823 million passengers (Jansen). However, this increase in demand puts greater pressure on airports to manage more flights with lower margins of time and error to properly guide airplanes out of their gates in an efficient manner. Many airline travelers have experienced the occasional nuisance of delayed flights and have become wary of the possibility that their upcoming flight may be delayed. Even with careful preparation, flight delays can surprise travelers as they occur on every airline at varying times of the day. Flight delays can cause increased stress on travelers and lead to lost time and energy.

Flight delays are not only a nuisance to passengers. Most commercial airplanes are on a set schedule with multiple flights planned throughout the country. A flight delay in one airport may have a compounding effect and generate delays for scheduled departures in other airports. If a delay is significant, the airline may be required to cancel future flights in the day and offer refunds to passengers, or pay for meals and overnight accommodations until a new flight can be offered (Mutzabaugh). I personally have faced this issue before and have had meals, hotel, and airport transfer covered by the airline when faced with a delay.

Airport terminal vendors also have good reason to pay attention to flight delays, albeit for a different reason. Passengers who are delayed can choose to either sit at their terminal gate or wander the terminal to alleviate boredom. With few options, many passengers opt to peruse the terminal restaurants and stores to make purchases while waiting for their delayed flight. Many terminal vendors have reported increased sales on days where the airport has experienced higher than usual delays (Zumbrun).

Many flight booking services recognize that delays are a serious concern for travelers. Some attempt to warn travelers if they believe a flight is relatively likely to be delayed. For example, Google Flights has built-in predictions that alert travelers if the flight they are about to book is expected to be delayed 30 minutes or longer. Unfortunately, Google and other companies do not release any details about the prediction methods they use. I decided to perform a time series analysis on flights departing from LAX Airport to see if any insight could be gained about predicting flight performance in the time and frequency domains. The results from this analysis could have implications for travelers, airline carriers, and terminal store vendors.

Data and Pre-processing

I obtained recent flight performance data from January 2014 through October 2016 from RITA, a division of the U.S. Bureau of Transportation Statistics. The data includes the number of minutes each flight was delayed, the departure airport, the date, and the airline carrier. When downloading the data I selected the option to only download flight data from LAX Airport. Once the data was downloaded, I decided to first convert the outcome variable into a binary indicator of whether the flight was delayed more than 30 minutes. I set 30 minutes as the cutoff because I felt that from personal experience this amount of time delay has led to significant frustration in the past. In addition, I wanted this work to be relevant to current industry precedents such as the 30 minute cutoff on Google Flights. There was a very small number of missing values and N/A values in the data which were removed.

After this round of cleaning, I decided to focus only on well-established airlines that have a large presence at LAX. To do this I eliminated all but the 10 most frequently used airlines in LAX during the study period. Once this was done, I then calculated the percentage of daily flights delayed over 30 minutes by dividing the sum of my binary variable by the total for each day. A percentage would be a more reliable indicator of performance than counts, because some days have more flights scheduled than others. Days with more delays may not necessarily translate to the overall performance of these airlines whereas a percentage of the total could.

Analysis – Time Domain

The first action taken after pre-processing was to plot the data over time. Figure 1 shows the percentage of daily delays during the entire collection period, except for the final two weeks that are held out of this analysis to assess prediction accuracy. The red line in this figure is a horizontal line denoting the average percent delayed over the study period. At a glance the data does not appear to be stationary as there are several periods of sustained above-average or below-average delays. In addition, we can see a few predictable outliers around the winter holidays of Christmas and Thanksgiving, days which routinely make news headlines for having airport delays.

Linear regression was run on this data to ascertain if there may be a general increasing or decreasing linear trend over time. The result of this test produced an insignificant slope and a P-Value of 0.35. As a result, I felt I did not need to detrend the data based on this linear model. While there was no linear trend, I did suspect a weekly cycle in the data.

I know from personal experience that flights tend to be busier on weekends, and less busy in the middle of the week. Based on this knowledge I decided to summarize my data by day of week. Figure 2 shows a bar plot of the results. Certain days of the week have on average higher delays than others. I decided to remove the weekly cycle to make the remaining data more stationary and suitable for time series analysis. Figure 3 shows the time series data with the weekly cycle removed. Analysis was then conducted on this form of the data.

To assess stationarity and begin to assess potential model fitting I plotted the ACF and PACF for the data which is shown in Figure 4. The ACF tails off relatively quickly, but does spike at lag 7 indicating that there is still a week-to-week relatedness in the data. The PACF decays quickly to insignificant or borderline insignificant values. The PACF of about 0.5 at lag 1 suggests a potential AR(1) process. It also indicates a relatedness of future days to the previous day. For example, it indicates that days with above average delays tend to be followed by days with a similar percentage of above average delays.

After reviewing the ACF and PACF, several attempts were made to fit an ARIMA model to the data. A couple of models with first differences were analyzed for comparison but every attempt was made to find a model without first differences as the interpretation of a first differenced model would not have a significant meaning. The model parameters, as well as AIC and BIC, are displayed in Table 1. The seasonal ARIMA (2,0,1)(1,0,0)[7] model had the lowest sum of AIC and BIC. As a result, this model was chosen for diagnostics and prediction of held out data.

The model diagnostics are shown in Figure 5. The standardized residuals of the model appear to be white noise and somewhat normally distributed. Some of the outliers on the positive side are quite high and give me some concern about the appropriateness of the model. The ACF of residuals is consistently below the threshold for significance and provides some confidence that the model is a good fit. The Ljung-Box statistics also lend more credence to this model as the p-values at all lags are above the significance threshold. The normal QQ plot serves to confirm my initial suspicions when eyeballing the standardized residuals. However, the QQ plot overall looks reasonably normal with only a few departures from normality from especially large outliers.

The next step taken in the analysis was to predict two weeks ahead and compare the results to the actual data that was held out. The plot of this prediction is displayed in Figure 6, where the light confidence band represents an 80% confidence interval and the darker band represents a 95% confidence interval. The predicted values are the blue line, and the actual data is in black. It is clear that the predicted values are not an exact match but the confidence intervals generally capture the prediction. Some of the actual values are quite close to the prediction during the first week, suggesting that this model may have some usefulness in day-to-day forecasting of flight delays.

Analysis – Frequency Domain

A frequency analysis was also conducted on this data. The first step taken was to plot a raw periodogram of the data which is shown in Figure 7. There is a strong peak corresponding to the weekly delay cycle, but there are lots of smaller peaks surrounding it which makes it difficult to determine if the peak is significant or not. I decided to take two smoothing approaches to attempt to find a better fit.

The first approach was to find AIC and BIC for various AR models. The plot of the AIC and BIC is displayed in Figure 8. It is clear from this graph that both AIC and BIC reach minimum values at lag 7, which indicates an AR(7) model may be a good choice to a smoothed periodogram. Figure 9 shows the AR(7) smoothed periodogram. Unfortunately, the smoothed periodogram does not show a clear peak and the maximum appears to be at or very close to zero indicating an unreasonably large periodicity in the data. The smoothing from AR(7) created a peak that is too broad to be meaningfully interpreted.

The second approach taken was smoothing by the Daniell Kernel. The smoothing parameters used were (2,0,2) as they bared some similarity to the best fitting ARMA model in the time domain analysis. The results of this smoothing are shown in figure 10. In this case the peak is much more well-defined and corresponds to the weekly periodicity.

Conclusion and Next Steps

The time domain analysis produced a seasonal ARMA model which seemed to fit the data reasonably well, but the variance of the predictions from this model produced confidence bands that were too large to be meaningful for distant forecasting. However, some of the predictions within the first few days seemed to be fairly accurate. This may serve as a starting tool for any parties which may be interested in day-to-day forecasting, such as airport logistics managers, or terminal vendors considering when to implement enticing sales promotions. However, the magnitude of the AIC and BIC in the model fitting, as well as some deviations from normality in the residual QQ Plot, leads to the belief that better models may be available. In the frequency domain, a clear weekly peak was identified even though the data had removed a weekly cycle. This result suggests there is a week-to-week relatedness in the data which was not captured when the weekly cycle was removed.

The initial results of this analysis are promising but there are many avenues for continued research. A good starting point for continued study would be to analyze potential covariates such as weather patterns. The LAX airport does not experience snow, but flights can often still become delayed due to adverse weather conditions such as heavy rain or fog. As an example, the maximum point in the data (July 18, 2015) was a day where a surprise rainstorm occurred in the middle of the summer after the area had experienced years of drought-like conditions. On this day nearly 50% of flights from LAX were delayed. It is also important to note that flights inbound to LAX may experience their own weather-related delays which affects later flights scheduled to depart from LAX to a new destination. A potential improved forecasting model may benefit from weather data at LAX, as well as weather data from the airport of the preceding flight inbound to LAX.

This data could also be analyzed for similarities to different times and different airports. This research only focused on LAX between 2014 and 2016. Another avenue of research may be to examine time series departure data from other major airports. Results could then be compared to the work done in this study to get a broader sense of how airlines experience delays around the country. It is also important to note that airline reporting policy has fluctuated over the last several decades. In recent years airlines have increased their estimated transit time between destinations (sometimes referred to as “schedule padding”) to either improve fuel economy, account for potential delays, or both (Morris). This could have an impact on how often flight delays occur over the years. If there are any patterns in schedule padding trends, a cut-point analysis might be able to identify them and provide more insight to travelers about the delays and travel times they should be prepared for. Researching these ideas may prove fruitful and lead to effective and helpful insights for travelers and business with a significant stake in the airline industry.

References

Jansen, Bart. *DOT: Airlines carried record 823 million passengers last year*. 17 March 2017. Web. 18 March 2017.

Morris, Hugh. *Why are flight times longer than they were 40 years ago?* 16 January 2017. Web. 18 March 2017.

Mutzabaugh, Ben. *Delayed flights: What is your airline responsible for?* 27 November 2013. Web. 18 March 2017.

Zumbrun, Joshua. *How Airports Profit From Your Wait*. 3 June 2008. Web. 18 March 2017.

Figures

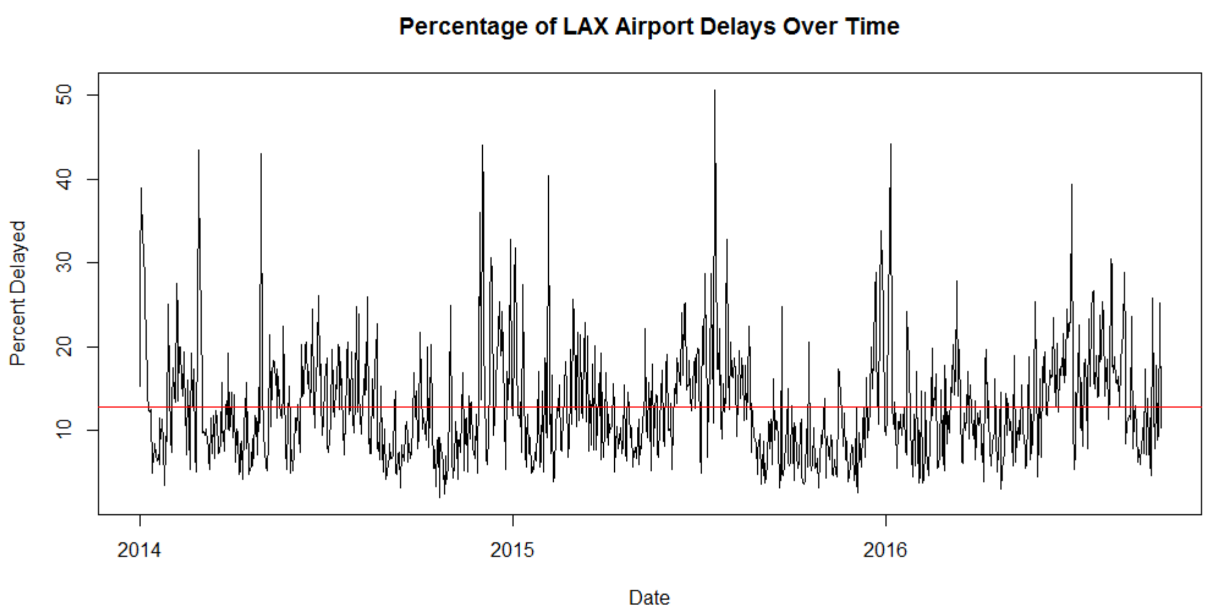


Figure 1 – Initial time series plot of the data with mean (red line)

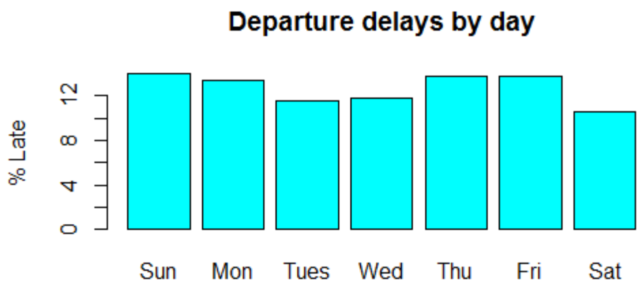


Figure 2 – Percentage of flight delays by day of week

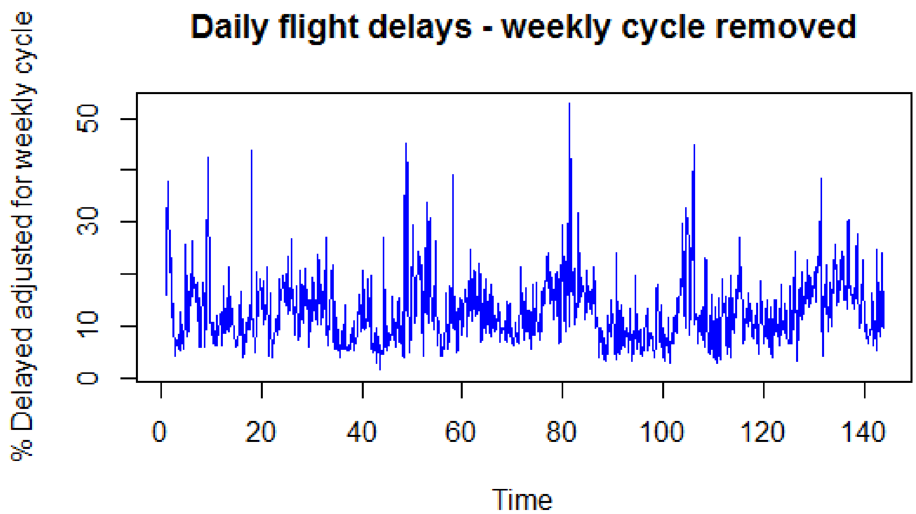


Figure 3 – Time series with weekly cycle removed

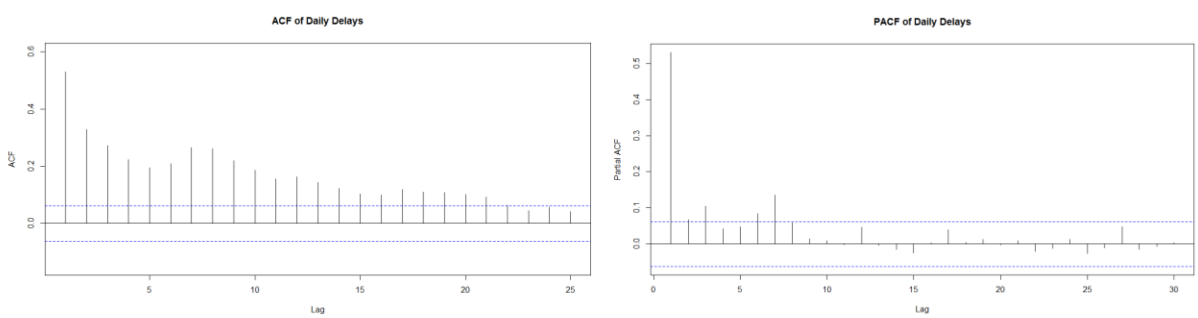


Figure 4 – ACF and PACF of daily delays data

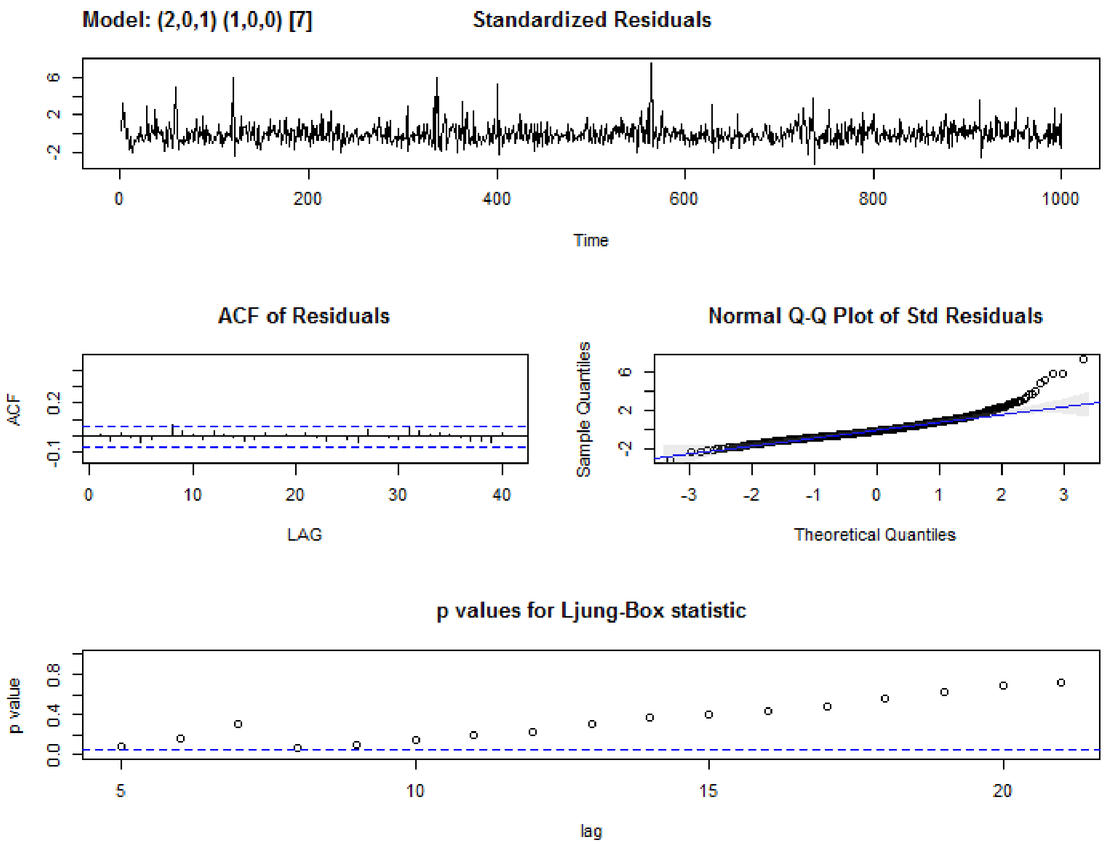


Figure 5 – model diagnostics for the seasonal ARIMA (2,0,1)(1,0,0)[7]

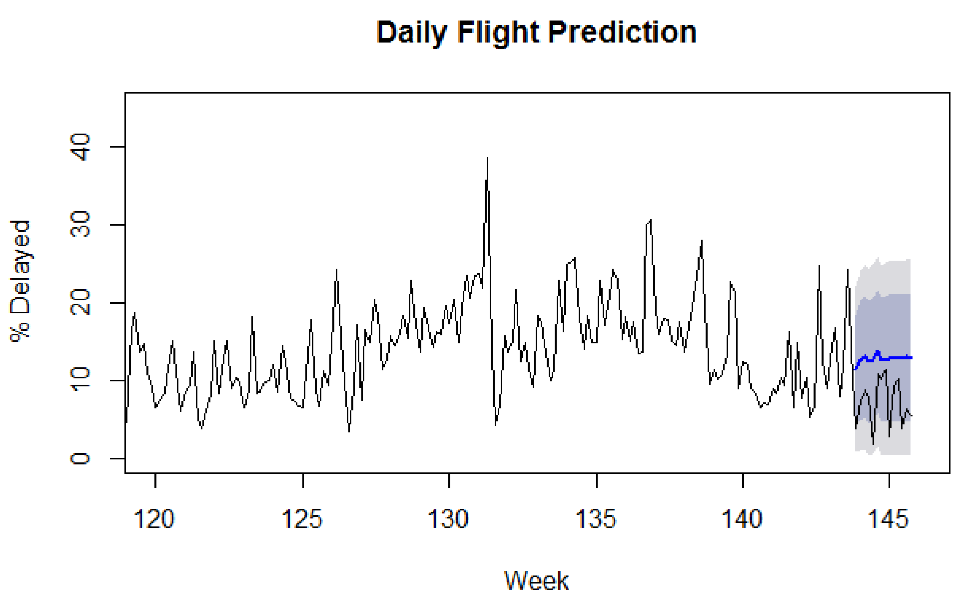


Figure 6 – Prediction on held out data in time series model

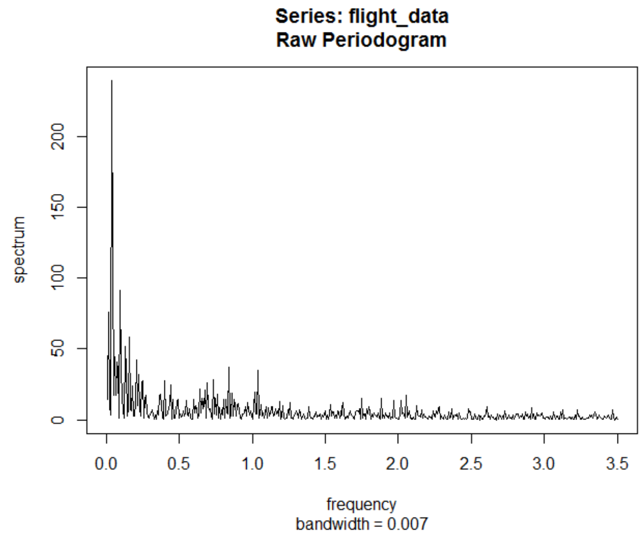
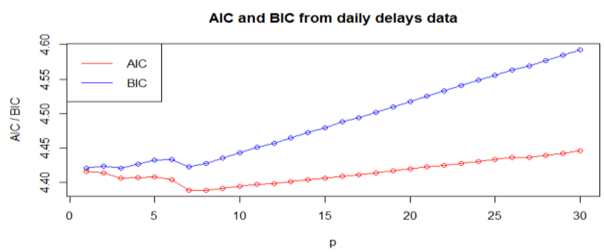
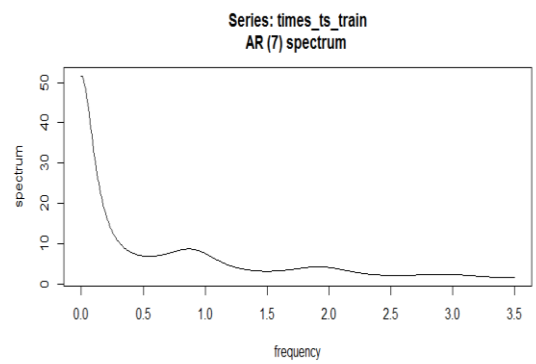


Figure 7 – Raw periodogram





Figures 8 and 9 – AIC/BIC plot and smoothed AR(7) periodogram

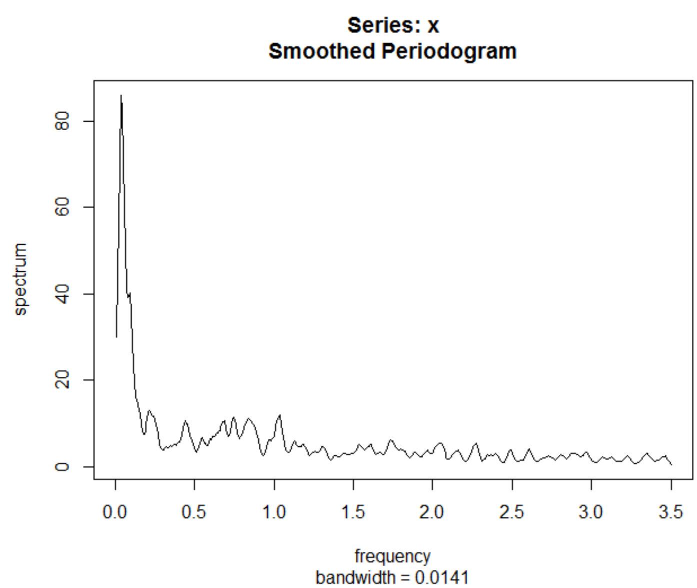


Figure 10 – Smoothed periodogram from (2,0,2) Daniell Kernel