Jordan Zoellmer
NEAS Time Series
Summer Session 2016
Student Project

An ARIMA Time Series Analysis of Observed Microsphere Brownian Motion

INTRODUCTION

Named after the 19th century botanist Robert Brown who first observed the phenomenon in 1827, Brownian motion describes the random movement of particles such as pollen grains or dust suspended in a gas or a liquid. After its formal discovery, debate ensued as to the cause of this random motion. It was not until almost 80 years later that Albert Einstein and Jean Perrin proved that the random movement of a pollen grain in water was the result of individual water molecules bombarding the larger particle in random and often unequal directions. Their discoveries confirmed the existence of atoms and molecules, naturally having major repercussions across many scientific fields.

As part of an undergrad contemporary physics course, I self-designed a lab in which I attempted to observe the Brownian motion of polystyrene (plastic) spheres in distilled water. I varied the size of the spheres as well as the viscosity of the fluid in order to measure certain properties relating to Brownian motion. However, after studying time series and autoregressive processes, I am intrigued at the potential to apply new statistical methods to analyze the same data from my undergraduate work. After all, Brownian motion is simply a natural realization of a random walk, a process that can be analyzed extensively using time series.

Theoretically, a random walk in one dimension would follow a first-order autoregressive process with $\phi = 1$. In other words, my next position is entirely determined by my current position and some white noise term that is independent of my current position. The model could be specified as:

$$Y_t = \phi Y_{t-1} + e_t = Y_{t-1} + e_t.$$

This equation would apply equally well in *either* the x-direction *or* the y-direction. Because I have both directions of data, I can analyze each independently as if they were a unique one-dimensional data set.

My primary goal from this report is to analyze how well the first-order autoregressive process fits the one-dimensional Brownian motion I observed in the undergrad lab. For example, the process should not be stationary; first-differences should be taken to obtain a stationary white noise process. I would like to confirm the existence of an underlying white noise process. I would also like to identify if any drift existed. If so, what is the average drift in both the x and y directions? Such analysis was difficult to pursue with no knowledge of time series, and drift plays a major factor in observational error of Brownian motion, potentially ruining the results and analysis. Through all of this, I hope to prove that the observed process does, in fact, follow random walk in both directions with a reasonable degree of uncertainty.

DATA

The data that will be used for this analysis comes from the observation of 3.0 micron diameter polstyrene spheres in distilled water. The spheres are viewed under a microscope for approximately five minutes, or until lost from view. Position readings using video capture technology are taken in 10 second intervals. We record position data like this for 12 spheres, hereafter labeled as spheres A-L. Data samples include time (t) readings, a horizontal (x) position, and a vertical (y) position. The third dimension is ignored due to inability to measure depth in a slide under a microscope. All of this data can be viewed in the excel data file, with each tab corresponding to an individual sphere being measured.

We will assume that all particles begin at a (0,0) origin at time $t = 0$. Three realizations of position observations if each sphere begins at the origin are displayed in the below figure. These are spheres A, B, and C. Similar graphs can be made for all 12 of the spheres in the data sample.

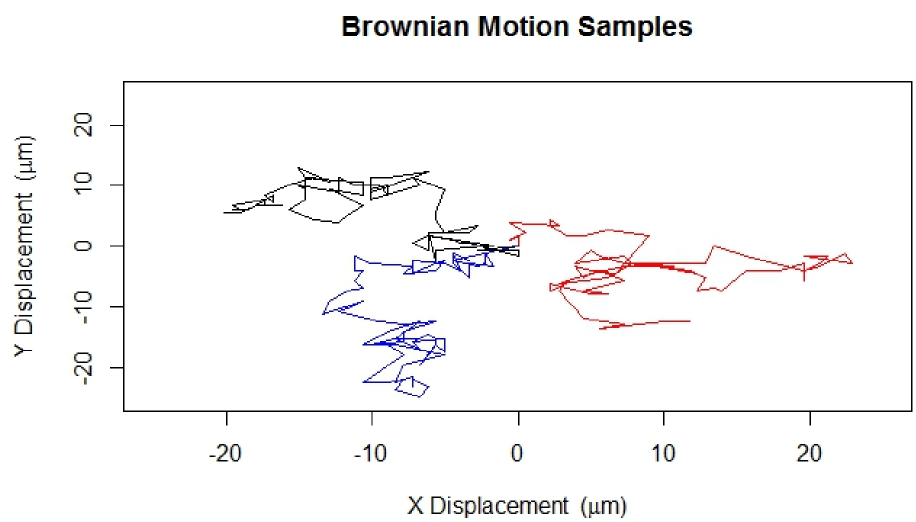

Figure 1: Three examples of the observed Brownian motion from this lab. These are spheres A, B, and C.

We will take the position data in the x and y directions and treat them independently. Each position at time t will be a value of the time series, either Y_t or X_t. We can similarly make a line plot like Cryer and Chan often do to depict the time series of each sphere for both Y_t and X_t. For sphere A, both time series are displayed below, again assuming the sphere begins at the origin. Both are a bit different, but both certainly appear to follow somewhat random walks to the naked eye. These plots can also be made for each of the other 11 spheres, and it is this data we will be analyzing. Essentially, we have 24 time series data sets, each of which could theoretically represent its own unique random walk. However, we will only analyze primarily the data for sphere A. But for data quality purposes it may be best to analyze as many as possible and take averages.

Of course, if these spheres do follow an AR(1) process, they will not be stationary. Ideally, we would model this by taking first-differences, as these should ideally follow a white noise process. As can be sen by a modification of the first equation, the first differences follow:

$$Y_t - Y_{t-1} = \nabla Y = e_t.$$

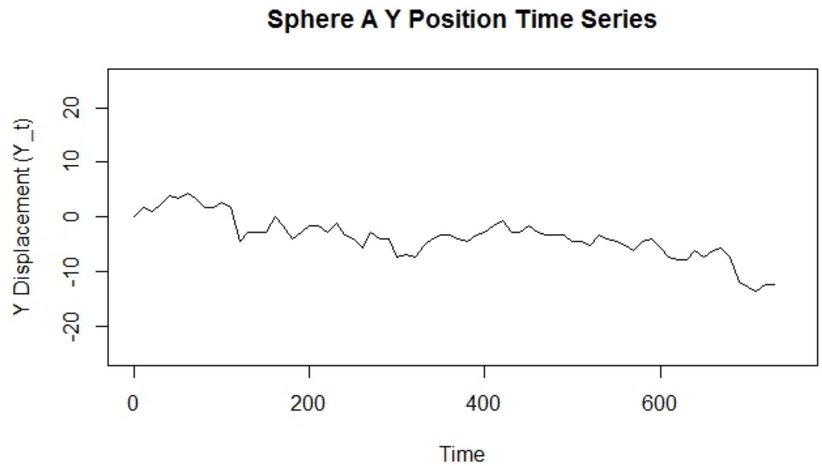

Figure 2: Time plot of Y motion for sphere A.

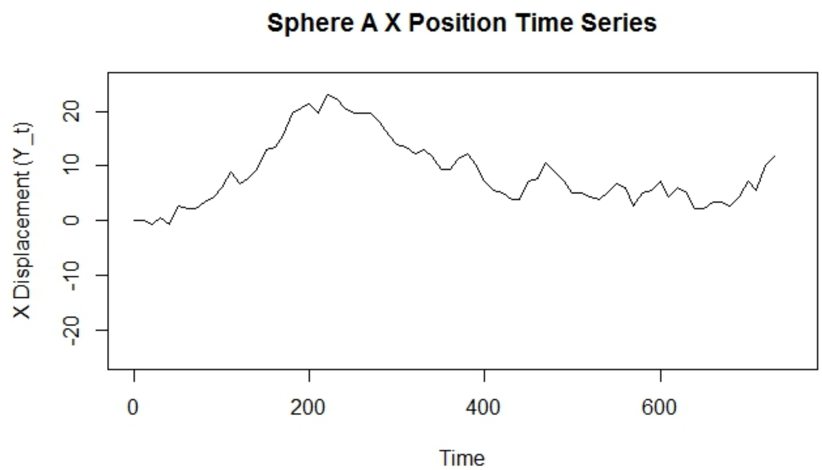

Figure 3: Time plot of X motion for sphere A.

Therefore, the data set included in the excel file also contains dt, dx, and dy, the first-differences of time and positions. Therefore, we also have 24 data sets of first differences that we can analyze to determine if a white noise process fits well. Thankfully, as these differences are independent of position, we can merge the entire set of dx and dy data into one large data set for analysis. This should lend plenty of credibility. If the first difference follow white noise, the process is most likely random walk. We will proceed to analyze the various components of the data and apply ARIMA processes to the data.

STATISTICAL TECHNIQUES

The first part of the analysis will be to examine the behavior of the time series of each of the 24 position observation series. From here on, we will use sphere A as an example, displaying the x and y analysis for sphere A. However, the analysis can easily be done on all 12 spheres by following the format in the R script. First, we analyze the autocorrelation of the time series. The lag 1 autocorrelation is evident by the strong positive correlation displayed in the next two graphs, when the Y_t is plotted against Y_{t-1} and X_t is plotted against X_{t-1}. Note that both directions exhibit

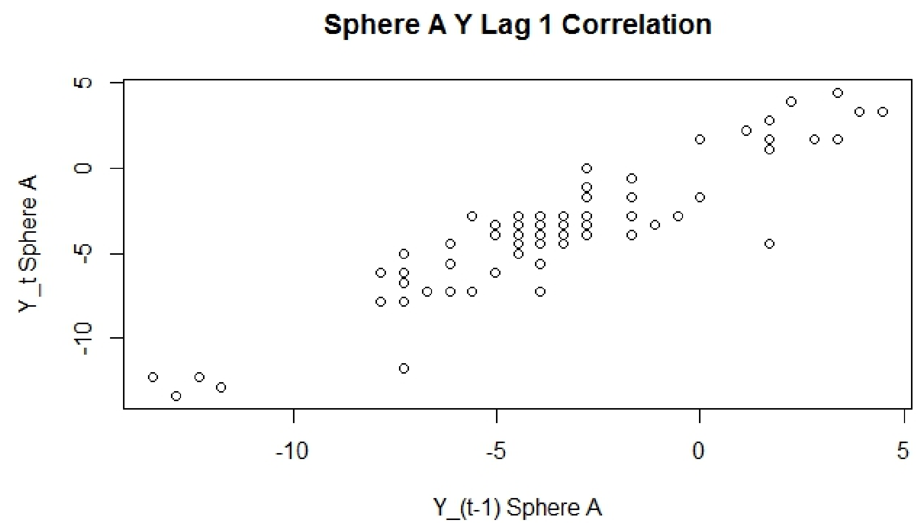

Figure 4: The lag 1 relationship for the Y position of sphere A.

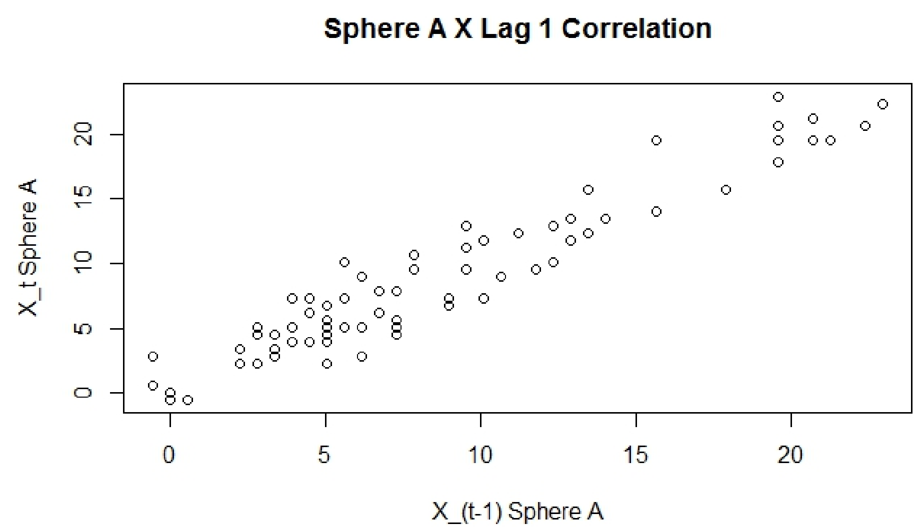

Figure 5: The lag 1 relationship for the X position readings of sphere A.

strong, positive correlation. In fact, we caclulate the correlation in the first graph for Y to be 0.915 and the correlation for the second graph of X positions to be 0.957. The relationship for an AR(1) model is that the lag 1 autocorrelation is given by $\rho_1 = \phi$. This would indicate that ϕ was over 0.9 for each. Of course, with any observations such as this the correlation will never be exactly 1. However, there seems to be strong evidence that the correlation does not differ significantly from 1. In other words, it seems likely that the time series is not stationary.

Next, we seek to explore other lags of autocorrelation. For a stationary AR(1) process we expect the autocorrelation function to decrease exponentially as the number of lags increase. The first nine lag plots are shown below in a grid for easy viewing for both X and Y positions. We note

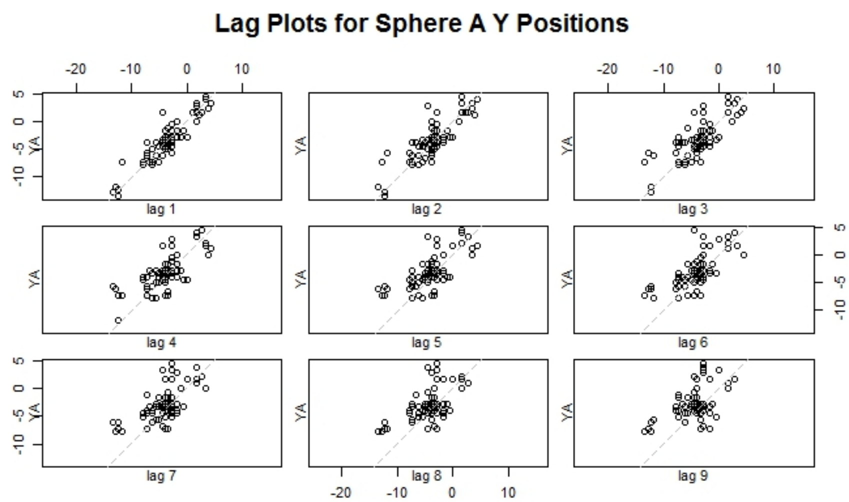

Figure 6: The first nine lags for the Y position of sphere A.

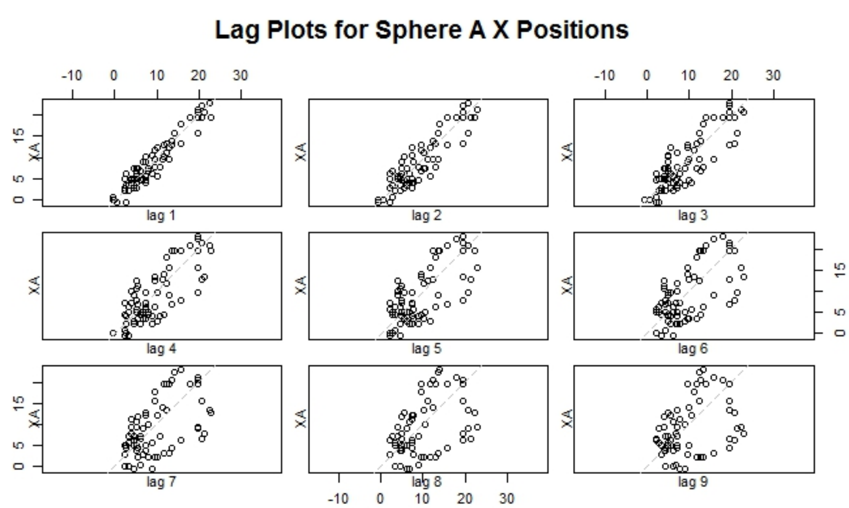

Figure 7: The first nine lags for the X position readings of sphere A.

that there is certainly some positive correlation persisting even into higher order lags. An initial reaction is most certainly that this is not a moving average process! Correlation remains far too high for high level lags for this to be a moving average process. Therefore, we are safe to conclude this is an AR process.

However, we can view the autocorrelation function a bit easier as often shown in the Cryer and Chan textbook. The autocorrelation functions for sphere A X and Y positions are displayed in the next two graphs. Neither plot is convincing of an exponential decrease in autocorrelation

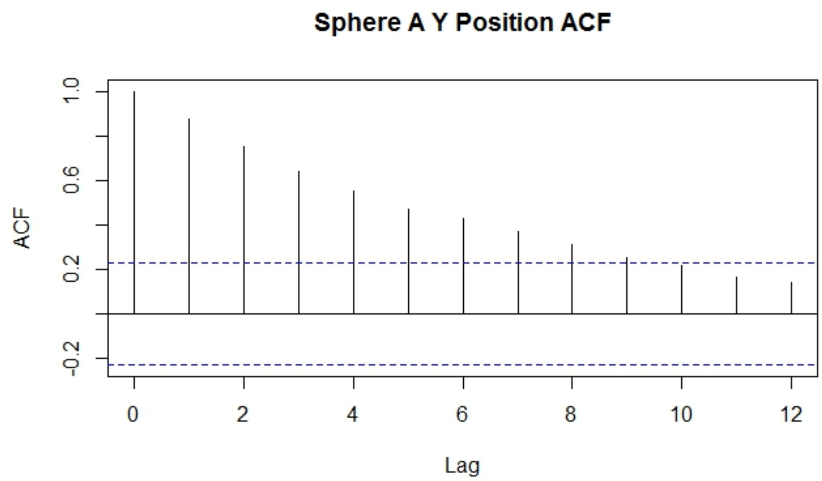

Figure 8: Autocorrelation function for the first 12 lags for the Y position for sphere A.

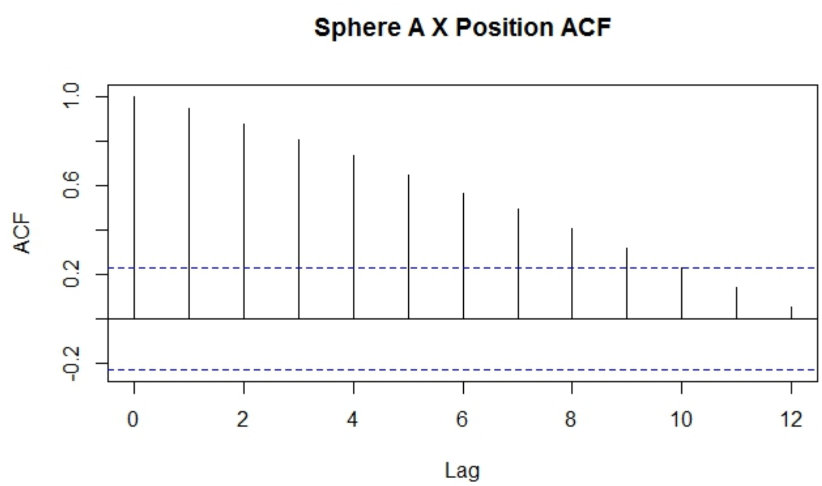

Figure 9: Autocorrelation function for the first 12 lags for the X position for sphere A.

over time. The graph provides further evidence of lack of stationary of this time series, making it seem possible that $\phi = 1$. At the very least, it will be worth taking first differences to see if the remaining process is indeed only white noise as would be the case if $\phi = 1$.

But first, we must ensure that the process is indeed AR(1) or ARI(1,1) as suspected, rather than a higher order AR process. This can be done by exploring the partial autocorrelation function. The PACF for an AR process acts much like the autocorrelation function for a moving average process. The PACFs for sphere A X and Y data are shown below. As we would expect of an

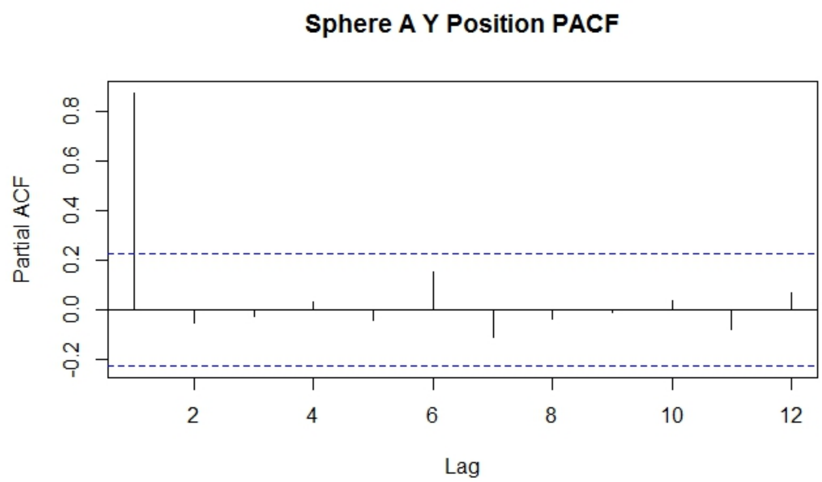

Figure 10: Partial autocorrelation function for the first 12 lags for the Y position for sphere A.

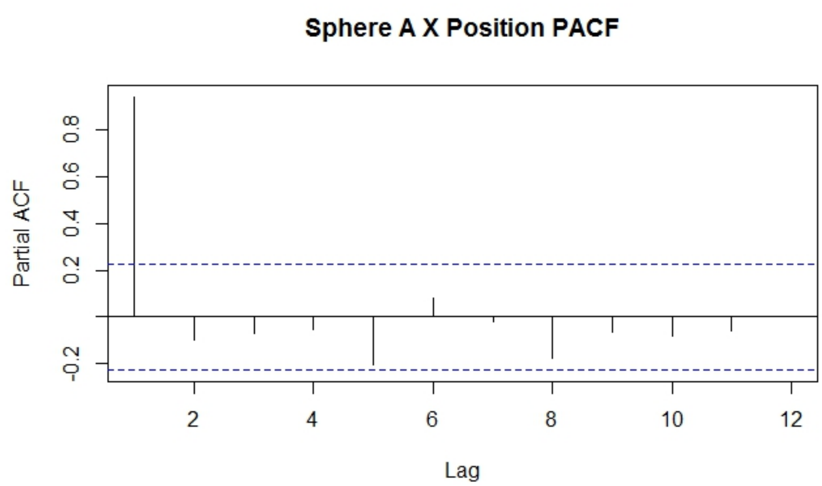

Figure 11: Partial autocorrelation function for the first 12 lags for the X position for sphere A.

AR(1) process, the PACF is significant for lag 1 in both graphs, but drops off dramatically and is not at all significant for the lags after 1. This confirms that the best model for this data is that of an AR(1), or maybe an ARI(1,1). This cannot be a moving average process, and this cannot be a higher level AR process.

Now we have established a very reasonable model for this data: the first-order autoregressive process with a high ϕ of around 0.915 for the Y data and 0.957 for the X data. This would be a reasonable position to take, and we could proceed from there to model the Brownian motion of the particles. Obviously these same techniques can be applied to the other 12 particles to obtain better estimates for the X and Y ϕ values. This would allow us to yield a more precise estimate with a tighter confidence interval. From there we can predict the movement of other particles we observe.

However, I am primarily interested if this model can be represented by the theoretically ideal random walk in which $\phi = 1$. I am still not convinced of the stationarity of the time series. In the next section, we take first differences and analyze the result. In particular, we will look at the time series for first differences and see if they behave like white noise. If so, we have evidence that the process is not stationary unless first differences are taken, in which case the observed Brownian motion is, in fact, a two dimensional random walk.

CORRECTIONS AND ADJUSTMENTS

In this section we look at the data for first difference of sphere A positions in both the X and Y directions. The first difference data is listed in the excel file and can be imported directly into R for analysis. We will first examine the time series plots of ∇Y and ∇X vs T. These two plots are displayed below. It is highly evident that neither process appears to drift from the mean of zero,

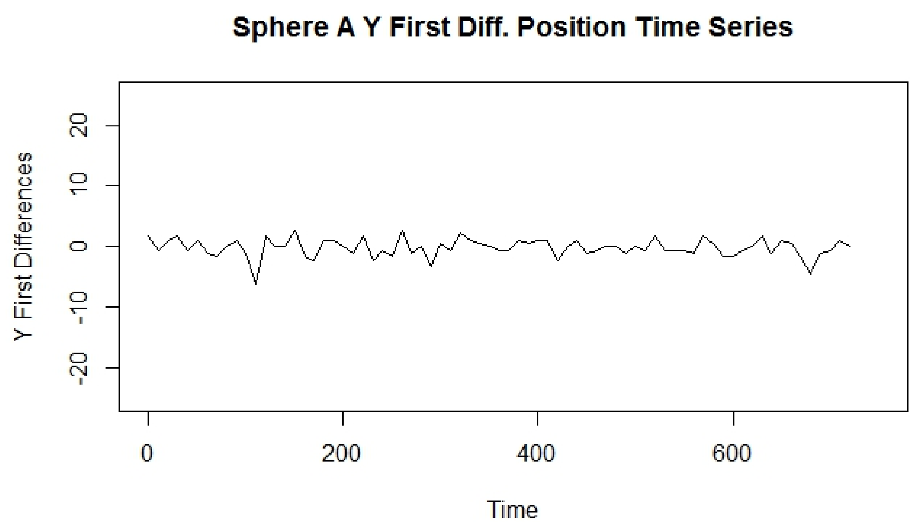

Figure 12: Time series plot of first differences in Y position for sphere A.

which is reassuring that drift is negligible in these observations. Furthermore, the process certainly appears to resemble white noise. Neither process has an apparent pattern to it. If this were, in fact, white noise, then the autocorrelation function would have no significant values for any lags greater than zero. We can check the autocorrelation functions of the first differences. These are displayed in the next two graphs, Figure 14 and Figure 15. We immediately notice that the no lags greater than zero have a significant autocorrelation function. This is definitive evidence that the first differences behave like white noise.

Two final graphs I would like to look at include first differences for all 12 spheres. Essentially

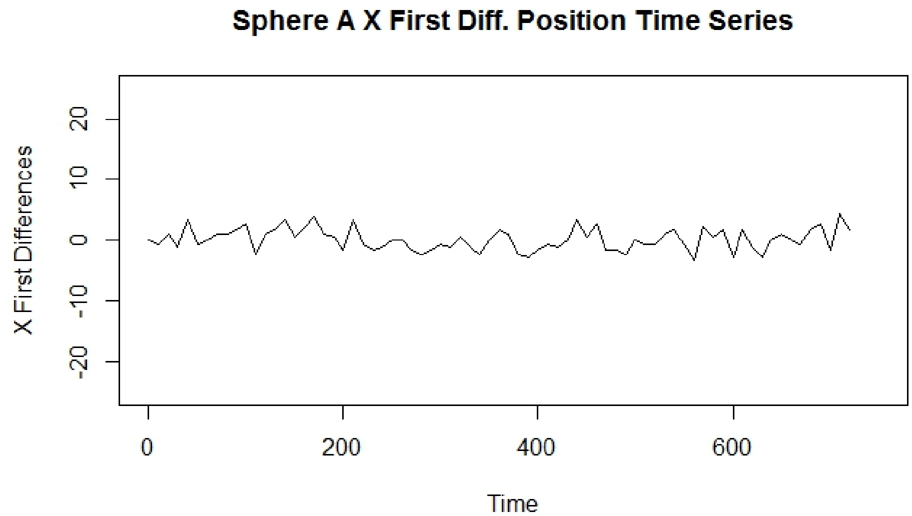

Figure 13: Time series plot of first differences in X position for sphere A.

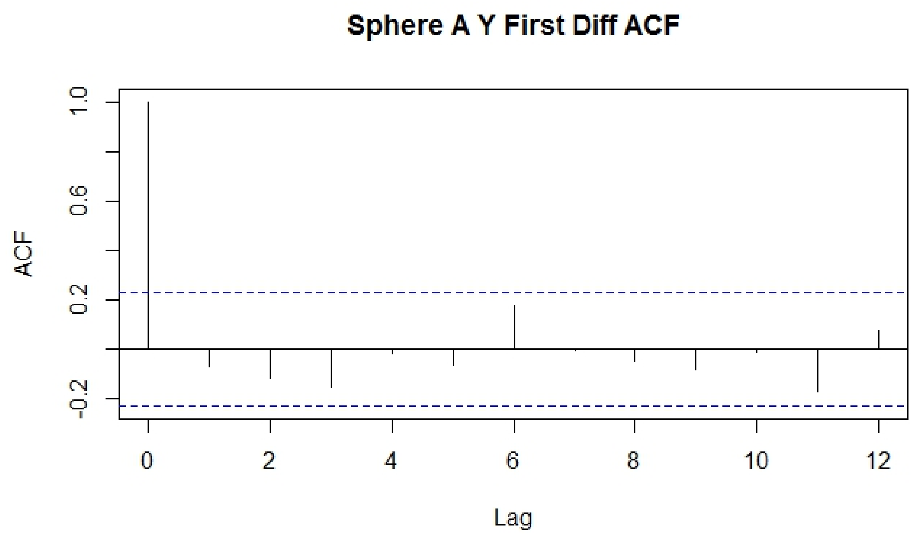

Figure 14: Autocorrelation function of the first differences of Y positions for sphere A.

we will be combining all 24 sets of first differences for X and Y from all spheres into one large data set. To prove beyond a doubt that white noise is a good model for the first differences, we can look at the normality of the first differences. First, we find that the mean of the first differences is -0.069 microns, with a variance of 2.93 and a standard deviation of 1.712. We can safely say that the mean can realistically be zero. If thefirst differences were in fact white noise, we would expect them to be approximately normally distributed. The following two plots display a histogram of the data against a normal curve and a normal qq plot. The histogram does look roughly normally

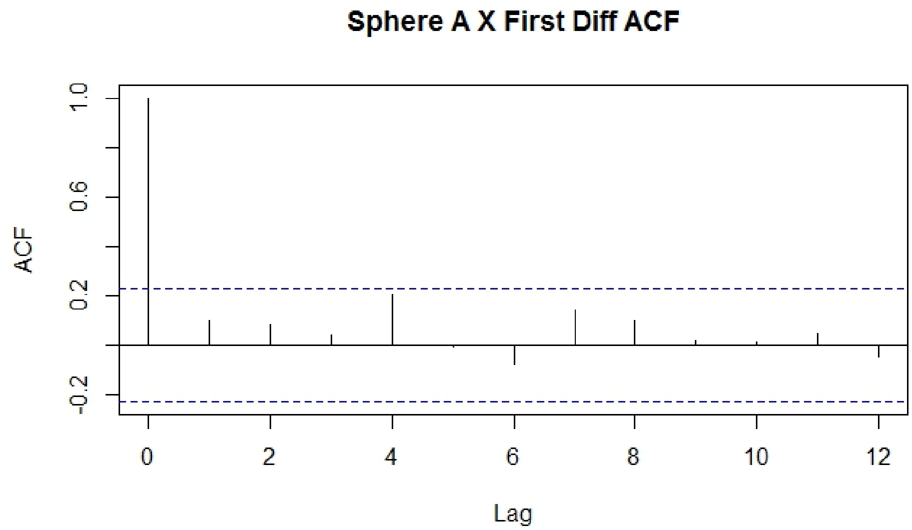

Figure 15: Autocorrelation function of the first differences of X positions for sphere A.

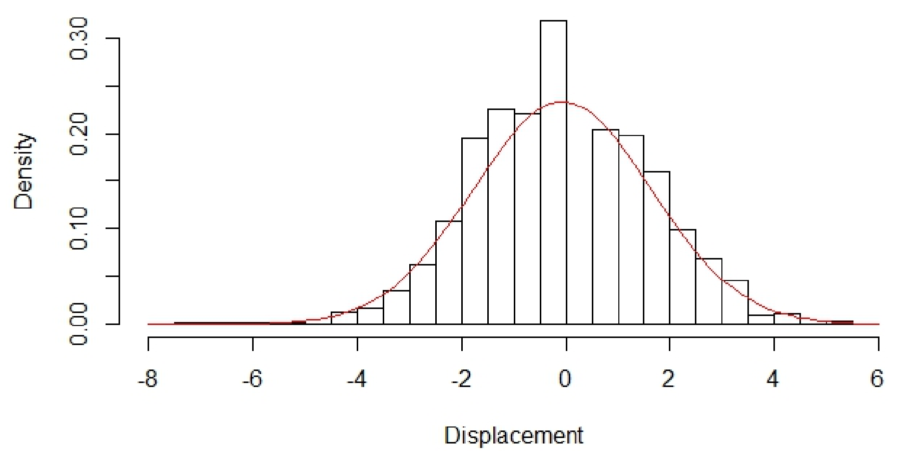

Figure 16: Histogram of the first difference of all measurements for all 12 spheres.

distributed, and the QQ plot confirms, with its linearity, that the differences are approximately normally distributed. These first differences can be reasonably modeled by white noise.

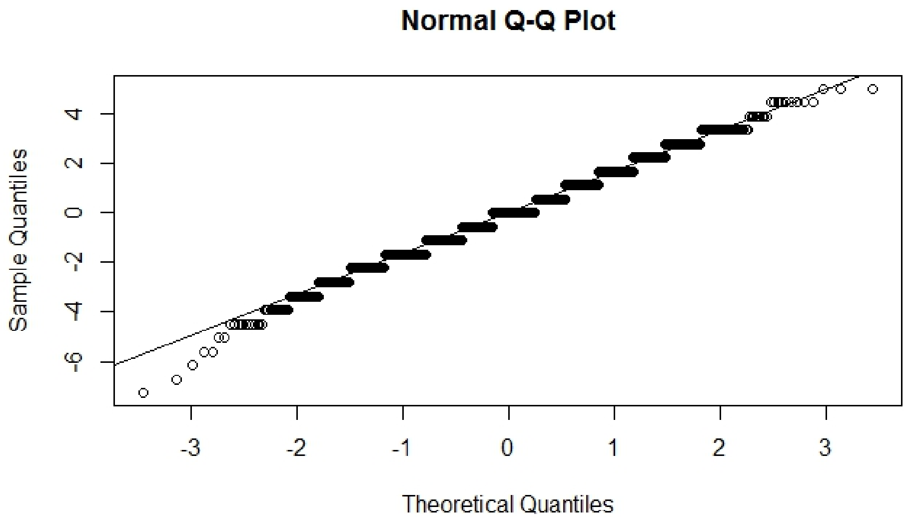

Figure 17: Normal QQ plot of first differences.

CONCLUSION

The goal of this analysis was to determine if the observed motion of polystyrene spheres in distilled water actually did follow Brownian motion as theorized. Brownian motion is essentially continuous time, multi-dimensional random walk. In this case, I observed the motion of 12 spheres in two dimensions. Analyzing just a single sphere in both dimensions, we were able to see that the autocorrelation resembled that of a first-order autoregressive model with a ϕ near 1. The partial autocorrelation analysis told us that higher order models were not valid. Theorizing that $\phi = 1$ and the process was not stationary, we looked at the first difference in an attempt to make it stationary. The resulting process resembled white noise, and we are therefore confident that the ideal model for the observed process is a random walk.

R SCRIPT

```
> # Jordan Zoellmer
> # Time Series VEE
> # Summer Session
> # July 11-Sept 2 2016
>
> # Student Project Data Analysis
>
> # This workbook is used for the Time Series Analysis
> # of the Brownian motion of 3 micro diameter spheres in water.
>
> # Import all 12 sphere observation data files.
> # The data comes from the tabs shown in the excel file.
> # Each tab was first put into its own csv file for importing.
> A.Data <- read.csv("~/Work/VEE Time Series Project/A Data.csv")
> B.Data <- read.csv("~/Work/VEE Time Series Project/B Data.csv")
> C.Data <- read.csv("~/Work/VEE Time Series Project/C Data.csv")
> D.Data <- read.csv("~/Work/VEE Time Series Project/D Data.csv")
> E.Data <- read.csv("~/Work/VEE Time Series Project/E Data.csv")
> F.Data <- read.csv("~/Work/VEE Time Series Project/F Data.csv")
> G.Data <- read.csv("~/Work/VEE Time Series Project/G Data.csv")
> H.Data <- read.csv("~/Work/VEE Time Series Project/H Data.csv")
> I.Data <- read.csv("~/Work/VEE Time Series Project/I Data.csv")
> J.Data <- read.csv("~/Work/VEE Time Series Project/J Data.csv")
> K.Data <- read.csv("~/Work/VEE Time Series Project/K Data.csv")
> L.Data <- read.csv("~/Work/VEE Time Series Project/L Data.csv")
> Cumulative.Displacements <- read.csv("~/Work/VEE Time Series Project/Cumulative Displacements
>
>
> # Plot 3 examples of Brownian motion in 2 dimensions.
> # Plotting spheres A, B, and C and their motion.
> XA = A.Data$Tx
> YA = A.Data$Ty
> plot(XA,YA,type="l", col="red", xlim = c(-25,25),ylim=c(-25,25), xlab =
+        expression(paste("X Displacement  (", mu, m, ")" )), ylab=
+        expression(paste("Y Displacement  ( ", mu, m, ")" )), main=
+        "Brownian Motion Samples")
>
> XB = B.Data$Tx
> YB = B.Data$Ty
> points(XB,YB,type="l", col="blue", add=T)
>
> XC = C.Data$Tx
> YC = C.Data$Ty
```

```
> points(XC,YC,type="l", add=T)
>
>
> # Plot time plot of time series for sphere A in both the X and the Y direction.
> TA = A.Data$T
> XA = A.Data$Tx
> YA = A.Data$Ty
> plot(TA,YA,type="l", col="black", xlim = c(0,750),ylim=c(-25,25), xlab = "Time",
+      ylab="Y Displacement (Y_t)", main="Sphere A Y Position Time Series")
> plot(TA,XA,type="l", col="black", xlim = c(0,750),ylim=c(-25,25), xlab = "Time",
+      ylab="X Displacement (X_t)", main="Sphere A X Position Time Series")
>
>
> # Look at lag 1 autocorrelation by comparing time series values
> # at t to at t-1.  All 24 graphs can be made with the same format as below for each table of
> YtA = A.Data$Ty[-1]  # exclude first term
> Yt_1A = A.Data$Ty[-74]   #exclude last term
> plot(Yt_1A,YtA, col="black", xlab = "Y_(t-1) Sphere A",
+      ylab="Y_t Sphere A", main="Sphere A Y Lag 1 Correlation")
> cor(YtA,Yt_1A)   # correlation is .9153
[1] 0.9152601
>
> XtA = A.Data$Tx[-1]  # exclude first term
> Xt_1A = A.Data$Tx[-74]   #exclude last term
> plot(Xt_1A,XtA, col="black", xlab = "X_(t-1) Sphere A",
+      ylab="X_t Sphere A", main="Sphere A X Lag 1 Correlation")
> cor(XtA,Xt_1A)   # correlation is .957
[1] 0.9569413
>
> # Lag plots of first nine lags for X and Y
> lag.plot(YA,lags = 9,do.lines=F, main = "Lag Plots for Sphere A Y Positions")
> lag.plot(XA,lags = 9,do.lines=F, main = "Lag Plots for Sphere A X Positions")
>
> # Autocorrelation functions for sphere A X and Y
> acf(YA, lag.max=12, plot = TRUE, main="Sphere A Y Position ACF")
> acf(XA, lag.max=12, plot = TRUE, main="Sphere A X Position ACF")
>
> # Partial acf for sphere A X and Y
> pacf(YA, lag.max=12, plot = TRUE, main="Sphere A Y Position PACF")
> pacf(XA, lag.max=12, plot = TRUE, main="Sphere A X Position PACF")
>
>
> # First Differences time series plots for sphere A X and Y
> dXA = A.Data$dx[-74]
> dYA = A.Data$dy[-74]
```

```
> TA = A.Data$T[-74]
> plot(TA,dYA,type="l", col="black", xlim = c(0,750),ylim=c(-25,25), xlab = "Time",
+      ylab="Y First Differences", main="Sphere A Y First Diff. Position Time Series")
> plot(TA,dXA,type="l", col="black", xlim = c(0,750),ylim=c(-25,25), xlab = "Time",
+      ylab="X First Differences", main="Sphere A X First Diff. Position Time Series")
>
> # ACFs of first differences
> acf(dYA, lag.max=12, plot = TRUE, main="Sphere A Y First Diff ACF")
> acf(dXA, lag.max=12, plot = TRUE, main="Sphere A X First Diff ACF")
>
>
> # Normality of first differences
> dx = Cumulative.Displacements$dx
> dy = Cumulative.Displacements$dy
> dz = c(dx,dy)    # merge x and y data sets
>
> mean(dz)  #-0.069
[1] -0.06920905
> var(dz)   #2.93
[1] 2.930115
> sd(dz)    #1.7117
[1] 1.711758
>
> hist(dz, breaks = seq(from = -8, to = 6, by = .5), freq=F, main="", xlab="Displacement")
> curve(dnorm(x,mean(dz),sd(dz)), add=T, col = "red")
> qqnorm(dz)
> qqline(dz)
```