Time Series Analysis: Modelling the Periodic Component of Monthly Data using Fourier Series Harmonic Analysis in R

ORVILLE D. HOMBREBUENO

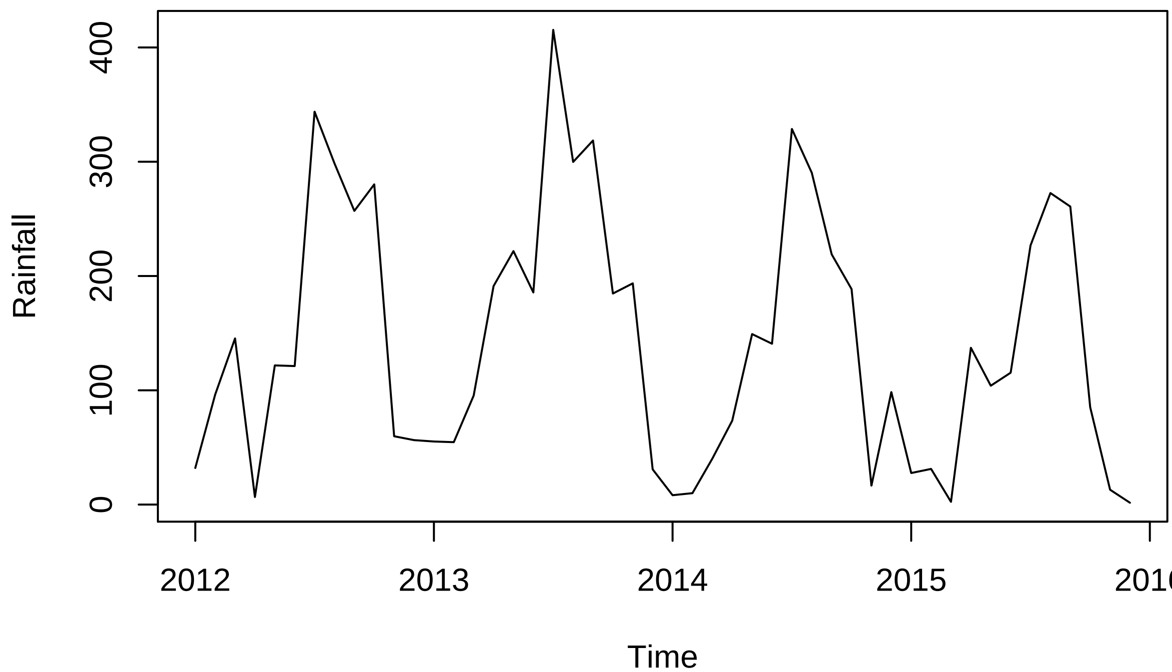
A time series is a sequence of observations taken sequentially in time. Many sets of data appear as time series: a monthly sequence of the quantity of goods shipped from a factory, a weekly series of the number of road accidents, daily rainfall amounts, hourly observations made on the yield of a chemical process, and so on. Examples of time series abound in such fields as economics, business, engineering, the natural sciences (especially geophysics and meteorology), and the social sciences. The main features of many time series are trends and seasonal variations that can be modelled deterministically with mathematical functions of time. Time series analysis is aimed at explaining the main features in the data using appropriate statistical models and descriptive methods.

In this vignette, I will be discussing time series analysis using basic Fourier series harmonic analysis (Weltner, John, Weber, Schuster, & Grosjean, 2014). I will show how to fit a Fourier series model to the periodic component of monthly rainfall data using R (R Core Team, 2019). In doing this, I will be following the model-building steps espoused by Box, Jenkins, Reinsel, & Ljung (2016): 1. model specification, 2. model fitting, 3. model diagnostics. I will not discuss theories behind Fourier series harmonic analysis nor will I discuss the importance of fitting a Fourier series model to monthly rainfall data. I will focus on the model fitting process and data analysis, for which I will be using R.

Model Specification

Our data is the monthly rainfall of Bayombong, Nueva Vizcaya from 2012 to 2015 in millimeters. Let us see the data.

```
> load("~/R/PhD/Advanced-Topics-in-Mathematics/Rainfall.RData")
> plot(Rainfall)
```

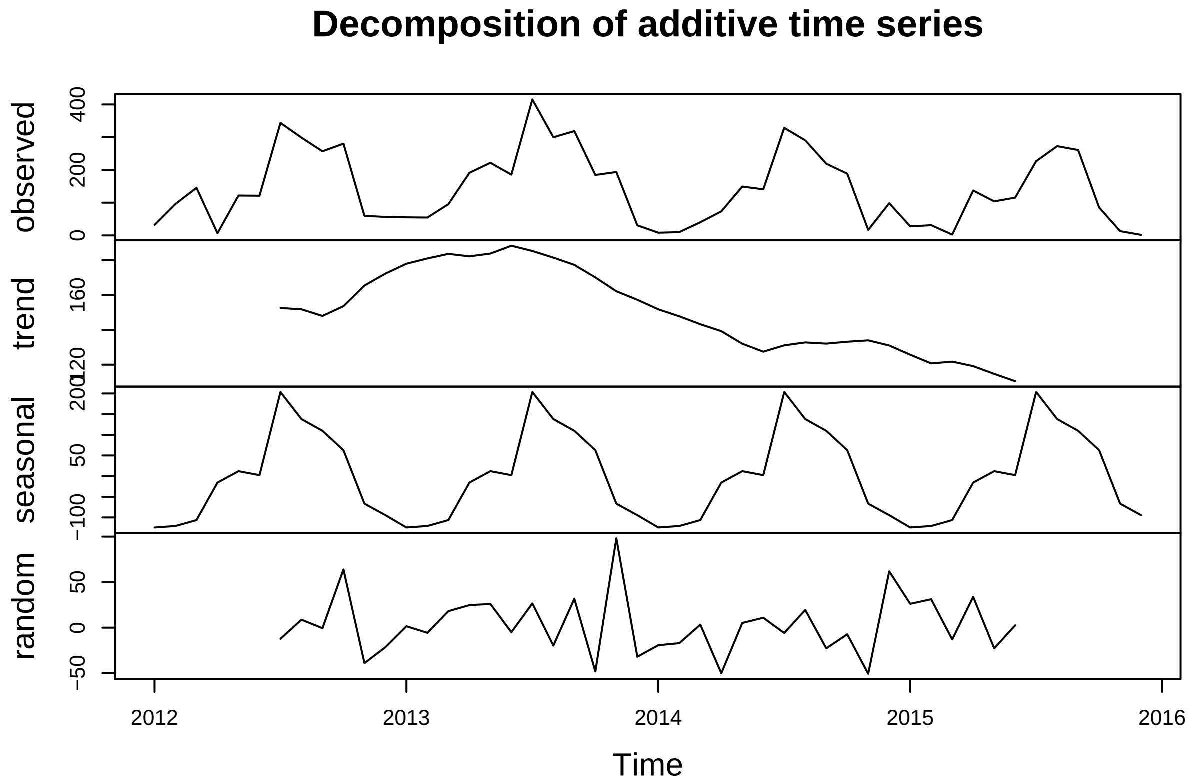


In model specification (or identification), an appropriate model for a given observed series is selected. To choose which model to use is not trivial. However, since my topic is Fourier series harmonic analysis, I will use Fourier series. Moreover, the data show periodic behavior. I have chosen this kind of data to make the use of Fourier series appropriate.

A given time series with respect to time $y(t)$ can be decomposed into the sum of its periodic component P_t, its trend component T_t, and random component ϵ_t:

$$y(t) = P_t + T_t + \epsilon_t. \tag{1}$$

```
> plot(decompose(Rainfall))
```



We will be modeling the seasonal or periodic component with Fourier series.

A function $f(x)$ of period $2L$ is repeated when x increases by $2L$, i.e.

$$f(x + 2L) = f(x) \tag{2}$$

If we put $z = \frac{\pi x}{L}$, then the function $f(z)$ is a periodic function of period 2π. As x increases from $-L$ to L, z increases from $-\pi$ to π. We have

$$f(z) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nz + b_n \sin nz). \tag{3}$$

Substituting z by $\frac{\pi x}{L}$,

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L}\right). \tag{4}$$

To solve for the coefficients we have the following equations:

$$
\begin{aligned}
a_0 &= \frac{1}{L}\int_{-L}^{L} f(x)\mathrm{d}x \\
a_n &= \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{n\pi}{L}x\mathrm{d}x \quad n = 1, 2\ldots \\
b_n &= \frac{1}{L}\int_{-L}^{L} f(x)\sin\frac{n\pi}{L}x\mathrm{d}x \quad n = 1, 2\ldots
\end{aligned} \tag{5}
$$

Defining our Fourier series for P_t we have

$$
P_t = a_0 + \sum_{n=1}^{\frac{P}{2}} \left(a_n \cos\frac{2\pi nt}{P} + b_n \sin\frac{2\pi nt}{P}\right) \tag{6}
$$

where a_0 is the mean monthly rainfall for the model, a_n and b_n are the Fourier coefficients, n is the number of harmonic, and P is the base period which is 12 months for 1 year. To make the equation more specific, we have

$$
\begin{aligned}
P_t =& a_0 + a_1\cos\frac{\pi t}{6} + b_1\sin\frac{\pi t}{6} + a_2\cos\frac{\pi t}{3} + b_2\sin\frac{\pi t}{3} + a_3\cos\frac{\pi t}{2} + b_3\sin\frac{\pi t}{2} \\
&+ a_4\cos\frac{2\pi t}{3} + b_4\sin\frac{2\pi t}{3} + a_5\cos\frac{5\pi t}{6} + b_5\sin\frac{5\pi t}{6} + a_6\cos\pi t.
\end{aligned} \tag{7}
$$

There are 6 harmonics in the equation and these harmonics corresponds to the periods: 12, 6, 4, 3, 2.4, and 2 months (Phien, Arbhabhirama, & Sunchindah, 1980). The sine term of the sixth harmonic is not included since it will be zero for all values of t.

```
> t <- 1:(4*12)
> s = c = matrix(nr = length(t), nc = 6)
> for(n in 1:6){c[,n] = cos(2*pi*n*t/12)
+ s[,n] = sin(2*pi*n*t/12)}
> s <- s[,-6]
```

We are going to perform backward elimination step wise time series regression to generate 10 Fourier series models with different number of harmonics. This will be discussed in the next section.

Model Fitting

Our models will inevitably involve one or more parameters whose values must be estimated from the observed series. Model fitting consists of finding the best possible estimates of those unknown parameters within a given model. In performing backward elimination step wise time series regression, we are going to use least squares for estimation which is default in the `lm()` function of R.

Model 1 is the model with all the terms. After fitting model 1 to the observed data, we will eliminate the term which has the highest p value or the least significant term.

```
> m1 <- lm(Rainfall ~ c + s)
> summary(m1)
```

```
Call:
lm(formula = Rainfall ~ c + s)

Residuals:
     Min       1Q   Median       3Q      Max 
-101.867  -27.725   -0.002   24.475  122.850 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  143.892      8.209  17.528  < 2e-16 ***
c1           -83.293     11.609  -7.175 1.94e-08 ***
c2           -28.339     11.609  -2.441   0.0197 *  
c3            10.972     11.609   0.945   0.3509    
c4           -13.334     11.609  -1.149   0.2583    
c5            25.355     11.609   2.184   0.0356 *  
c6            -8.453      8.209  -1.030   0.3100    
s1           -97.351     11.609  -8.386 5.49e-10 ***
s2            27.573     11.609   2.375   0.0230 *  
s3            -4.411     11.609  -0.380   0.7062    
s4            12.688     11.609   1.093   0.2817    
s5            -3.560     11.609  -0.307   0.7609    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 56.87 on 36 degrees of freedom
Multiple R-squared:  0.7987,    Adjusted R-squared:  0.7372 
F-statistic: 12.99 on 11 and 36 DF,  p-value: 1.815e-09
```

We can see in the model summary that the 5th harmonic of the sine term has the highest p value among the terms. Model 2 will then be the model without this term.

```
> m2 <- lm(Rainfall ~ s[,1] + s[,2] + s[,3] + s[,4] + c[,1] + c[,2] + c[,3]
+          + c[,4] + c[,5] + c[,6])
> #s[,5] is not included
> summary(m2)
```

```
Call:
lm(formula = Rainfall ~ s[, 1] + s[, 2] + s[, 3] + s[, 4] + c[,
    1] + c[, 2] + c[, 3] + c[, 4] + c[, 5] + c[, 6])

Residuals:
     Min       1Q   Median       3Q      Max
-102.551  -26.508   -2.431   21.410  124.630

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  143.892      8.108  17.747  < 2e-16 ***
s[, 1]       -97.351     11.466  -8.490 3.25e-10 ***
s[, 2]        27.573     11.466   2.405   0.0213 *
s[, 3]        -4.411     11.466  -0.385   0.7027
s[, 4]        12.688     11.466   1.107   0.2756
c[, 1]       -83.293     11.466  -7.264 1.26e-08 ***
c[, 2]       -28.339     11.466  -2.471   0.0182 *
c[, 3]        10.972     11.466   0.957   0.3448
c[, 4]       -13.334     11.466  -1.163   0.2523
c[, 5]        25.355     11.466   2.211   0.0333 *
c[, 6]        -8.453      8.108  -1.043   0.3039
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 56.17 on 37 degrees of freedom
Multiple R-squared:  0.7982,    Adjusted R-squared:  0.7437
F-statistic: 14.64 on 10 and 37 DF,  p-value: 4.765e-10
```

Now we can see that the 3rd harmonic of the sine term is the least significant term in model 2. This term will be gone in model 3. This goes on until 10 models are generated.

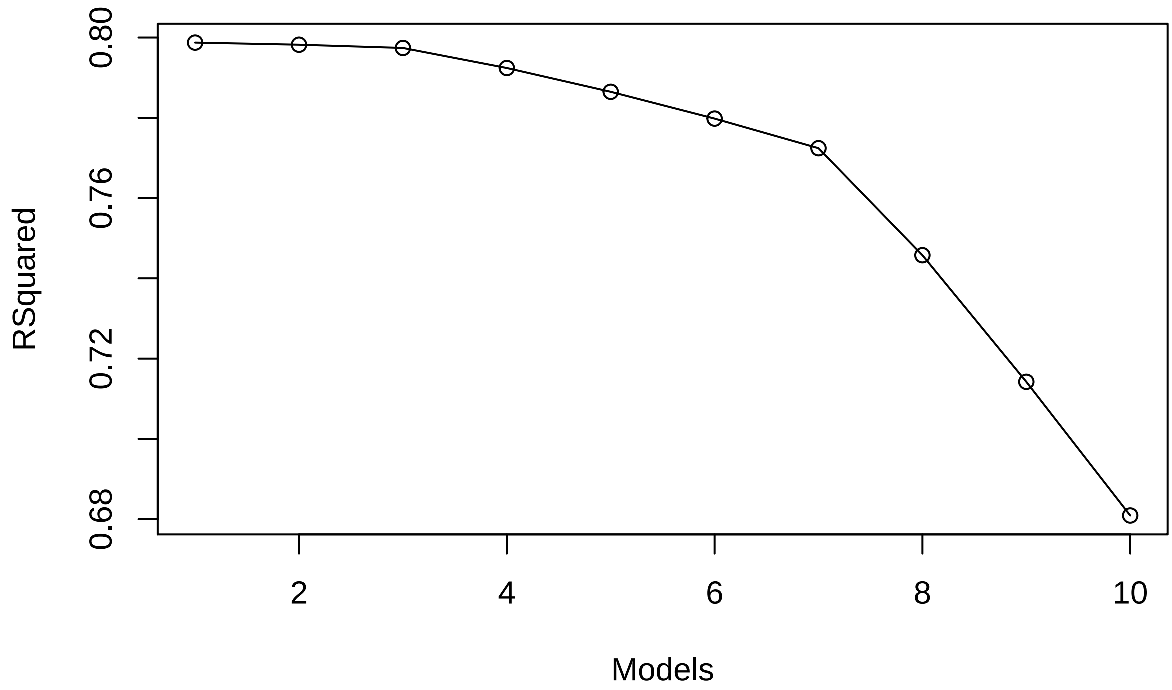
```
> m3 <- lm(Rainfall ~ s[,1] + s[,2] + s[,4] + c[,1] + c[,2] + c[,3] + c[,4]
+          + c[,5] + c[,6])
> #s[,5], s[,3] are not included
> m4 <- lm(Rainfall ~ s[,1] + s[,2] + s[,4] + c[,1] + c[,2] + c[,4] + c[,5]
+          + c[,6])
> #s[,5], s[,3], c[,3] are not included
```

```r
> m5 <- lm(Rainfall ~ s[,1] + s[,2] + s[,4] + c[,1] + c[,2] + c[,4] + c[,5])
> #s[,5], s[,3], c[,3], c[,6] are not included
> m6 <- lm(Rainfall ~ s[,1] + s[,2] + c[,1] + c[,2] + c[,4] + c[,5])
> #s[,5], s[,3], c[,3], c[,6], s[,4] are not included
> m7<-lm(Rainfall ~ s[,1] + s[,2] + c[,1] + c[,2] + c[,5])
> #s[,5], s[,3], c[,3], c[,6], s[,4], c[,4] are not included
> m8 <- lm(Rainfall ~ s[,1] + s[,2] + c[,1] +c[,2])
> #s[,5], s[,3], c[,3], c[,6], s[,4], c[,4], c[,5] are not included
> m9 <- lm(Rainfall ~ s[,1] + c[,1] + c[,2])
> #s[,5], s[,3], c[,3], c[,6], s[,4], c[,4], c[,5], s[,2] are not included
> m10 <- lm(Rainfall ~ s[,1] + c[,1])
> #s[,5], s[,3], c[,3], c[,6], s[,4], c[,4], c[,5], s[,2], c[,2]
> #are not included
```

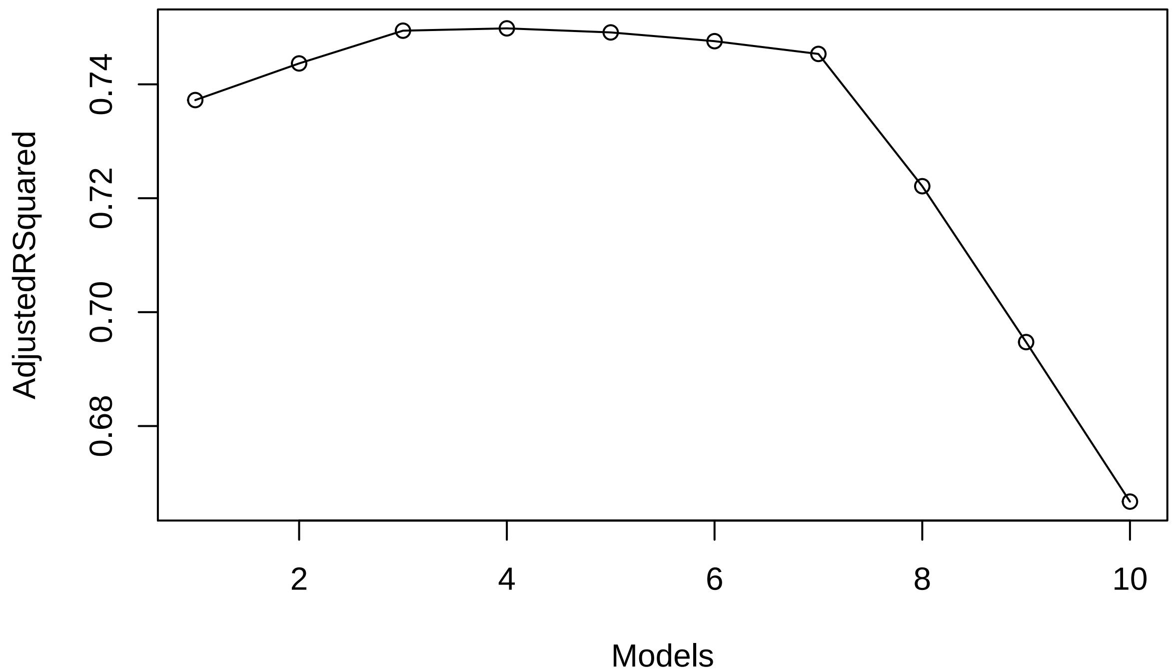
Model Diagnostics

Model diagnostics is concerned with assessing the quality of the model that we have specified and estimated. Here we are going to see which is the best fit model. We are going to compute the coefficient of determination R^2, adjusted coefficient of determination $\bar{R}^2$, residual standard error (RSE), and Akaike Information Criterion (AIC) of each model and then compare them to determine the best-fit model for the periodic component of the observed series.

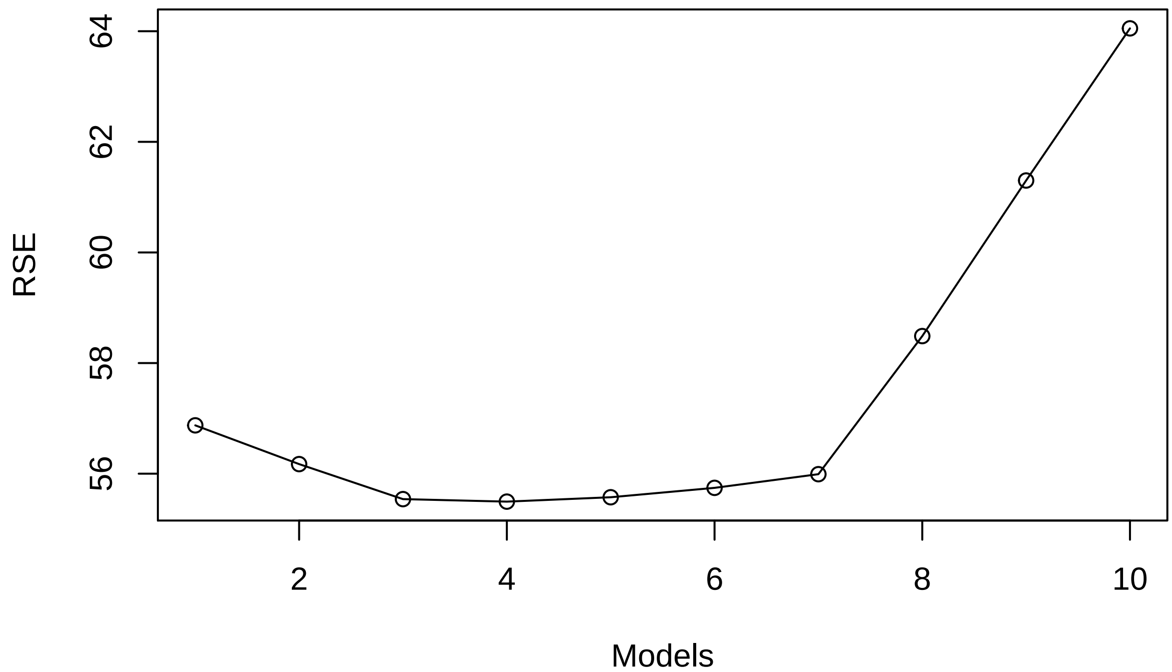
```
> sm1 <- summary(m1)
> sm2 <- summary(m2)
> sm3 <- summary(m3)
> sm4 <- summary(m4)
> sm5 <- summary(m5)
> sm6 <- summary(m6)
> sm7 <- summary(m7)
> sm8 <- summary(m8)
> sm9 <- summary(m9)
> sm10 <- summary(m10)
>
> RSquared <- c(sm1$r.squared, sm2$r.squared, sm3$r.squared, sm4$r.squared,
+               sm5$r.squared, sm6$r.squared, sm7$r.squared, sm8$r.squared,
+               sm9$r.squared, sm10$r.squared)
>
> Models <- 1:10
> plot(Models, RSquared, type = "o")
```



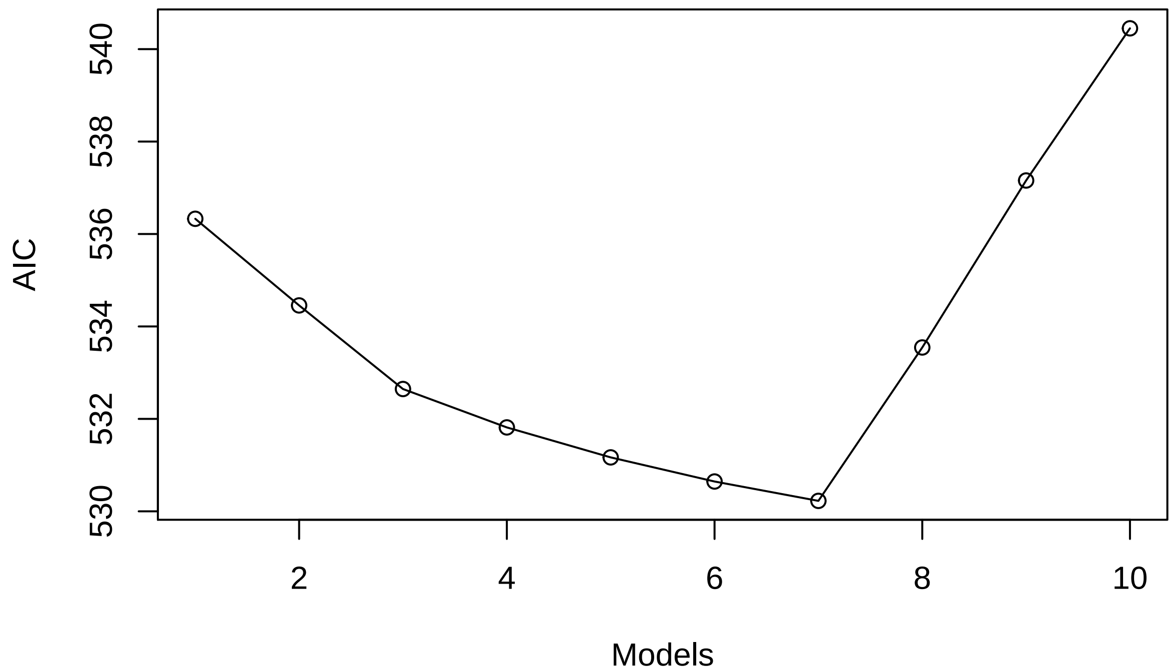

```
> AdjustedRSquared <- c(sm1$adj.r.squared, sm2$adj.r.squared,
+                       sm3$adj.r.squared, sm4$adj.r.squared,
+                       sm5$adj.r.squared, sm6$adj.r.squared,
+                       sm7$adj.r.squared, sm8$adj.r.squared,
+                       sm9$adj.r.squared, sm10$adj.r.squared)
>
> plot(Models, AdjustedRSquared, type = "o")
```




```
> RSE <- c(sm1$sigma, sm2$sigma, sm3$sigma, sm4$sigma, sm5$sigma,
+          sm6$sigma, sm7$sigma, sm8$sigma, sm9$sigma, sm10$sigma)
>
> plot(Models, RSE, type = "o")
```

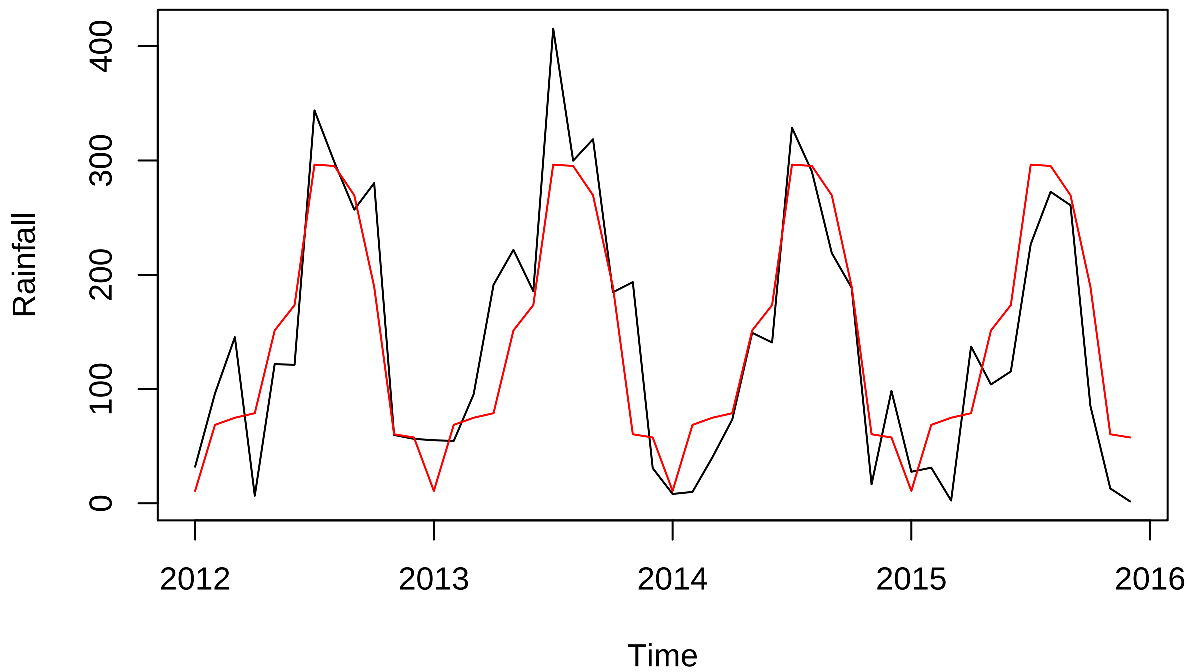



```
> A <- AIC(m1,m2,m3,m4,m5,m6,m7,m8,m9,m10)
> AIC <- A$AIC
> plot(Models, AIC, type = "o")
```



Looking at the results, there is a sudden increase in the change of values for R^2, $\bar{R}^2$, RSE, and AIC from model 7 to model 8. This tells us that model 7 is the best-fit model. The figure below is the observed series together with model 7.

```
> fv <- ts(m7$fitted.values, frequency = 12, start = 2012)
> ts.plot(Rainfall, fv, gpars = list(col = c("black", "red")),
+         ylab = "Rainfall")
```




```
> summary(m7)
```

```
Call:
lm(formula = Rainfall ~ s[, 1] + s[, 2] + c[, 1] + c[, 2] + c[,
    5])

Residuals:
     Min       1Q   Median       3Q      Max
-104.321  -39.054   -3.743   28.728  133.173

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  143.892      8.082  17.805  < 2e-16 ***
s[, 1]       -97.351     11.429  -8.518 1.07e-10 ***
s[, 2]        27.573     11.429   2.413   0.0203 *
c[, 1]       -83.293     11.429  -7.288 5.65e-09 ***
c[, 2]       -28.339     11.429  -2.480   0.0173 *
c[, 5]        25.355     11.429   2.218   0.0320 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 55.99 on 42 degrees of freedom
Multiple R-squared:  0.7724,    Adjusted R-squared:  0.7453
F-statistic: 28.51 on 5 and 42 DF,  p-value: 1.727e-12
```

Looking at the model summary, the best fit Fourier series model is:

$$\begin{aligned} P_t =&143.892 - 83.293\cos\frac{\pi t}{6} - 97.351\sin\frac{\pi t}{6} - 28.339\cos\frac{\pi t}{3} + 27.573\sin\frac{\pi t}{3} \\ &+ 25.355\cos\frac{5\pi t}{6}. \end{aligned} \tag{8}$$

The best-fit model has 6 terms: the mean term, the first harmonic, the second harmonic and the cosine term of the fifth harmonic – accounting to the 12-, 6-, and 2.4-month periods respectively – all of which are significant having p-values lesser than 0.05.

Bibliography

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (5th ed.). Hoboken, New Jersey: John Wiley & Sons Inc.

Phien, H. N., Arbhabhirama, A., & Sunchindah, A. (1980). Distribution of monthly rainfall in northeast thailand. *Southeast Asian Studies*, 110–123.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

RStudio Team. (2019). *RStudio: Integrated development environment for r*. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Weltner, K., John, S., Weber, W. J., Schuster, P., & Grosjean, J. (2014). *Mathematics for physicist and engineers: Fundamentals and interactive study guide* (2nd ed.). Springer-Verlag Heidelberg.