Methodological Climate Model
12/02/2025

Climate science models can be complicated and confusing largely because the models are the products of many “ready-made” packages. However, these packages hind the fundamentals deep in the convoluted workflow of “model training and testing”, as if the models themselves were indeed blackboxes. This article intends to open up the “blackboxes” and offer an easy-to-follow approach to climate science modeling. The example used is the historical sea level at Key West, FL.

The methodological approach is to divide the model into three components: trend, seasonality, and autoregression. The components are additive so that one modeling step is used for each component and the final model is the sum of the three components. In fact, each step decomposes the response variable data into the component model plus the residual term which is then modeled with the next component plus residual term until all three component models are complete.

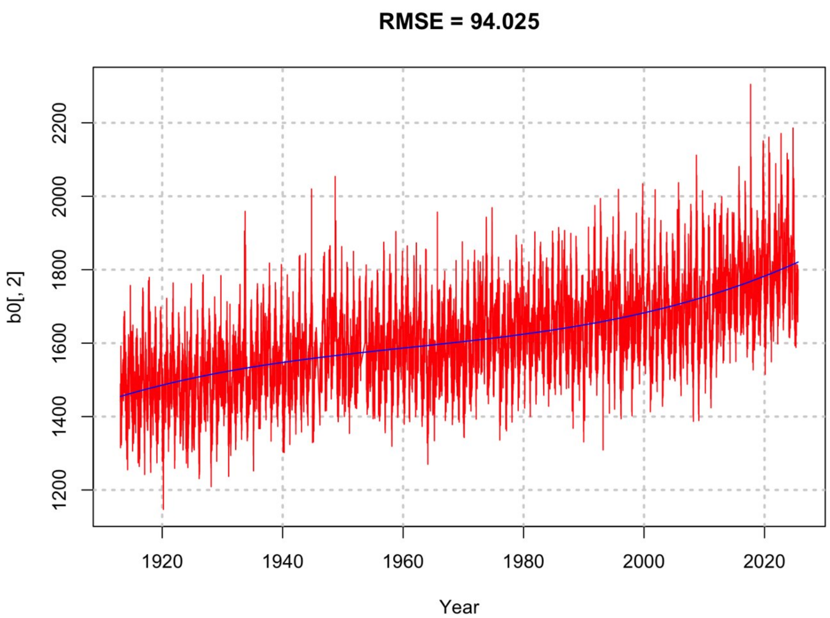

Figure 1, Sea Level Measurements at Key West, FL.

Trend

Figure 1 shows cubic linear model fitting the Key West data (data file d242.csv). A higher order linear model does not improve the root mean square error (RMSE). The trend model is done with function “trend()”.

```
trend=function(t,y){m=lm(y~t+I(t^2)+I(t^3)); return(m$fitted.values)}
```

Seasonality

A general approach to modeling seasonality is to use the Fourier transformation and ensure the model includes all important frequency components. Figure 2 shows the spectrogram of the residual term from the trend model.

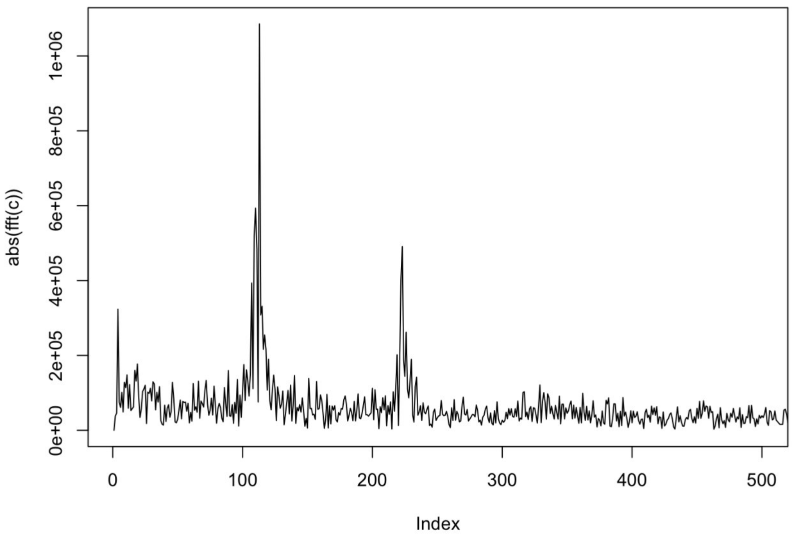

Figure 2, Fourier transformation of the residual term from the trend model.

Several frequency components locate at the frequency index of 110 and 220 (40214/110 = 365 days). As such, the seasonality model must include all the frequency component below 250. Figure 3 shows the modeling results. The seasonality model is accomplished by function season():

```
season=function(y,pe=pe){ftr=fft(y);for (i in pe:length(ftr)) {ftr[i]=0};
       iff=Re(fft(ftr,inverse=T))/length(ftr); return(2*iff)}
```

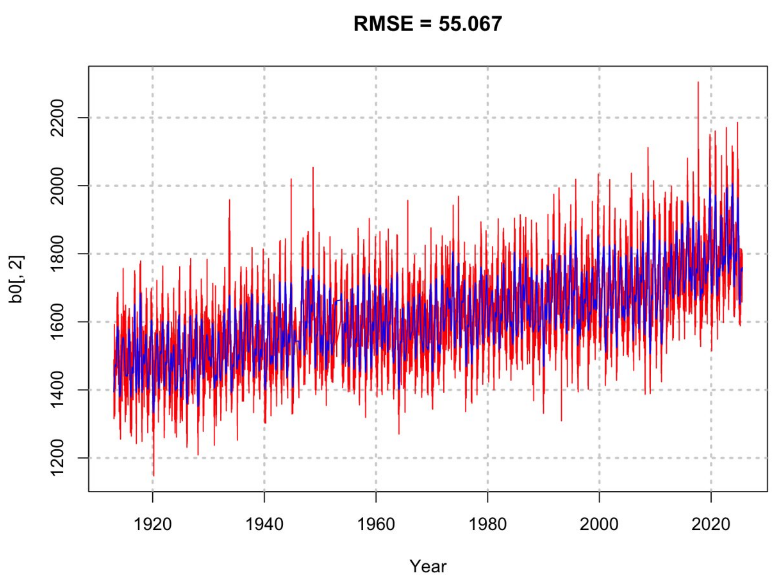

Figure 3, Sea Level Measurements at Key West, FL and Blue Line Model of both the Trend and Seasonality.

where parameter pe is the number of frequency components in the model. It is clear that the seasonality model reduces the RMSE from 94.025 to 55.067, a significant improvement.

Autoregression

While the trend model only uses time as the “explanatory variable”, the seasonality model uses frequencies as the explanatory variables. Autoregression, on the other hand, treats some previous response variable as the explanatory variables. A “lag” operator generates the corresponding variables whose number is determined using the partial autocorrelation function (PACF) of the residual term (d) from the seasonality model as shown in Figure 4. Five lagged terms are used.

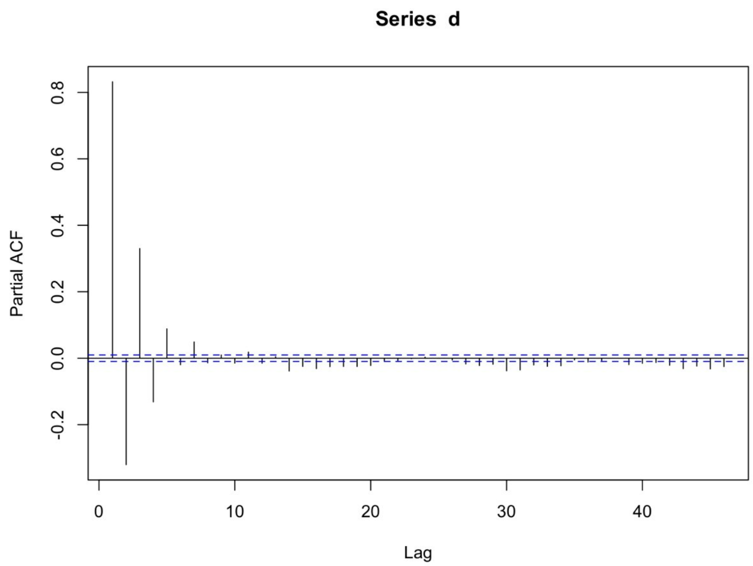

Figure 4, Partial Autocorrelation Function of the Residual Term (d) from the Seasonality Model.

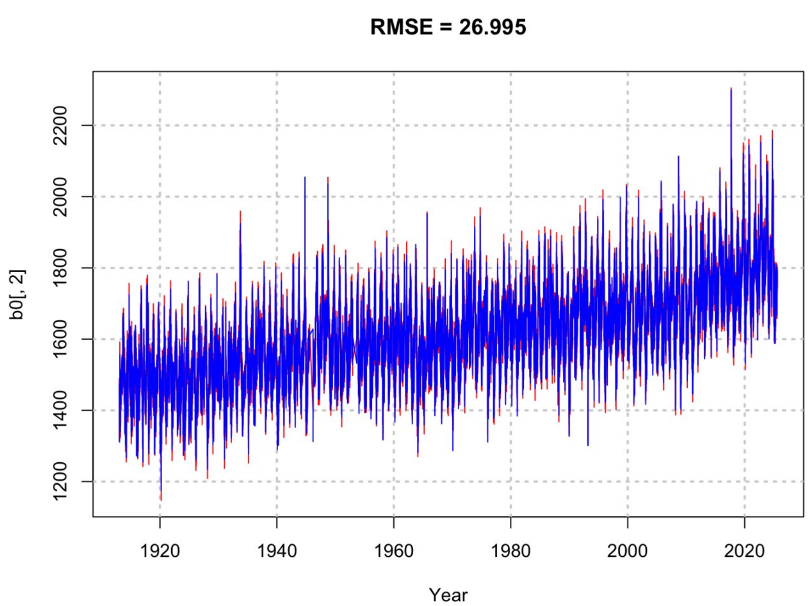

Figure 5, Combined Model with Trend, Seasonality, and Autoregression Components.

The lag operator and autoregression are made using lag(), auto() and autor() functions.

```
lag=function(x,n=1){c(rep(NA,n),x[-((length(x)-n+1):length(x))])}
auto=function(x,n){y=matrix(x,ncol=1);for(i in 1:n)
{y=cbind(y,lag(y[,i]))};return(as.data.frame(y))}
autor=function(x){m=lm(V1~. ,data=x); return(m$fitted.values)}
```

The combined model is shown in Figure 5. The RMSE or SD of the combined model is reduced to 26.995 from 55.067. A closeup view of the model is also in Figure 6.

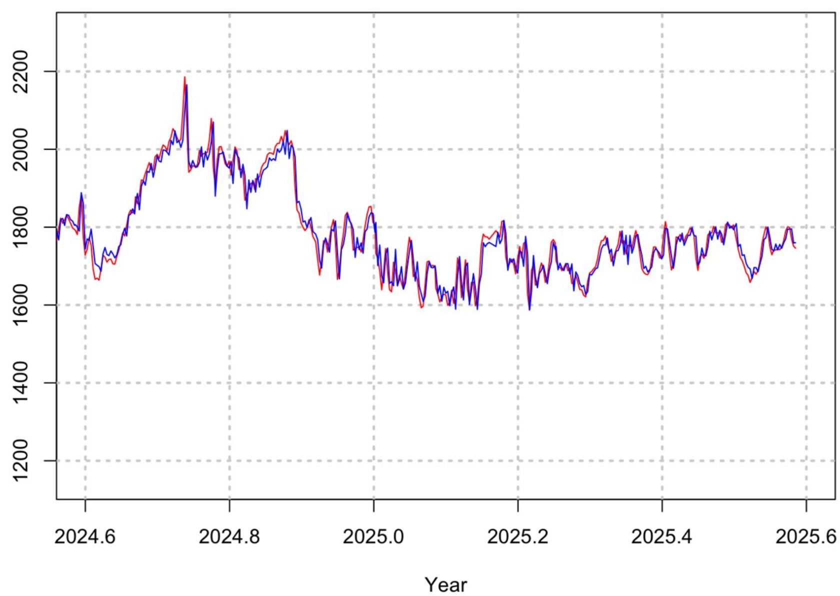

Figure 6, A Closeup View of Figure 5 of Actual Measurements (red line) and Model Prediction (blue line)

Concluding Remarks

The historical data from Key West, FL show that the sea level there follows approximately a 120 year cycle as the trend from 1910 to 2025 (Figure 1). A semi-annual cycle (183 days) plus an annual cycle (365 days) form frequent changes in the sea level (Figure 2). The even shorter term sea level changes (less than 6 months) are not attributed to the general trend nor the seasonality of the phenomenon, rather driven by a hypothetical “internal state” of the phenomenon that can be described by the previous sea levels as the proxy of the internal state. The optimal number of previous daily sea levels that serves as the proxy is 5 days as presented in this article (Figure 4).

Appendix R Script

```
pe=250; n=5; year=1900
input=function(fn){y=read.table(fn,header=F,sep=",");data.frame(date=y[,1]+(y[,2]-1)/
12+y[,3]/365,level=y[,4])}
lag=function(x, n=1){c(rep(NA, n),x[-((length(x)-n+1):length(x))])}
trend=function(t,y){m=lm(y~t+I(t^2)+I(t^3)); return(m$fitted.values)}
auto=function(x,n){y=matrix(x,ncol=1);for(i in 1:n)
{y=cbind(y,lag(y[,i]))};return(as.data.frame(y))}
autor=function(x){m=lm(V1~. ,data=x); return(m$fitted.values)}
season=function(y,pe=pe){ftr=fft(y);for (i in pe:length(ftr)) {ftr[i]=0};
       iff=Re(fft(ftr,inverse=T))/length(ftr); return(2*iff)}
dec=function(y){return(sprintf("%.3f",y))}

b1=input("d242.csv");
b1=b1[b1[,2]>0,];
b0=b1[b1[,1]>year,]
yr=b0[,1];

## trending
t=trend(yr,b0[,2])
c=b0[,2]-t
plot(yr,c, type="l"); grid(lty=2)
plot(b0[,1],b0[,2], type="n", xlab="Year",main=paste("RMSE =",dec(sd(c)))); grid(lwd=2)
lines(b0[,1],b0[,2],col="red")
lines(yr,t,col="blue")

## seasonality
plot(abs(fft(c)), type="l", xlim=c(1,500))
s=season(c,pe)
d=c-s
plot(yr,d, type="l"); grid(lty=2)
plot(b0[,1],b0[,2], type="n", xlab="Year",main=paste("RMSE =",dec(sd(d)))); grid(lwd=2)
lines(b0[,1],b0[,2],col="red")
lines(yr,t+s,col="blue")

## autocorrelation
pacf(d)
a=autor(auto(d,n))
s=s[-(1:n)]; t=t[-(1:n)]; yr=yr[-(1:n)]; d=d[-(1:n)]
h=d-a;
plot(yr,h, type="l"); grid(lty=2)

## combined
f=t+s+a
plot(b0[,1],b0[,2], type="n", xlab="Year", main=paste("RMSE =",dec(sd(h)))); grid(lwd=2)
lines(b0[,1],b0[,2],col="red")
lines(yr,f,col="blue")
```