Assignment1

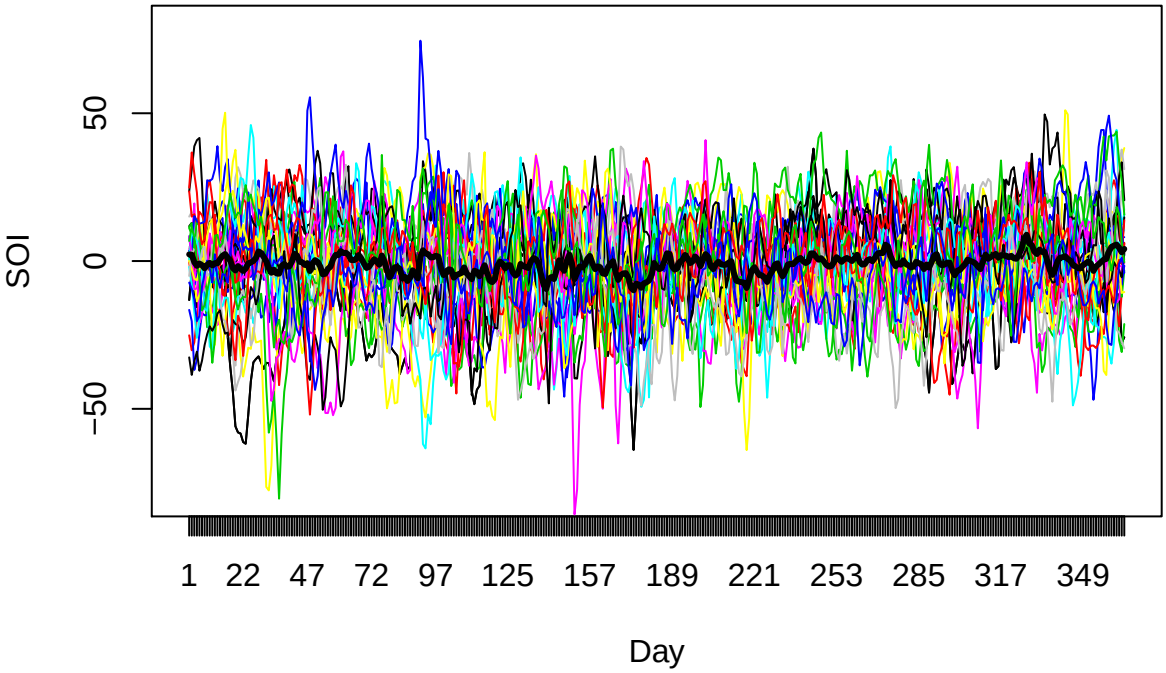
Tiffany Woodley

5/25/2020

Question 5: Functional Data Analysis

The data used for this question comes from 2 datasets, first being a data set containing daily values of the Southern Oscillation Index(SOI). The second dataset contains monthly rainfall values from an airport in Bloemfontein. The aim of this question is to see whether annual rainfall could be modelled from the daily SOI values of the respective year.

The data is first reconstructed to have each year's daily SOI values as a row (observation). The SOI measurements were taken on each day of the year resulting in 365 predictor variables. Years 1991 and 2020 did not contain measurements for all of the days and so were removed, leaving 28 years (observations). The plot below shows the 28 observations with the mean value in black.



The annual rainfall was calculated using the measired monthly rainfall at the bloemfoentein airport. This was done by summing over the summer months (January, February, March, October, November and December) .

In order to predict the annual rainfall from the SOI values we need to evaluate the model

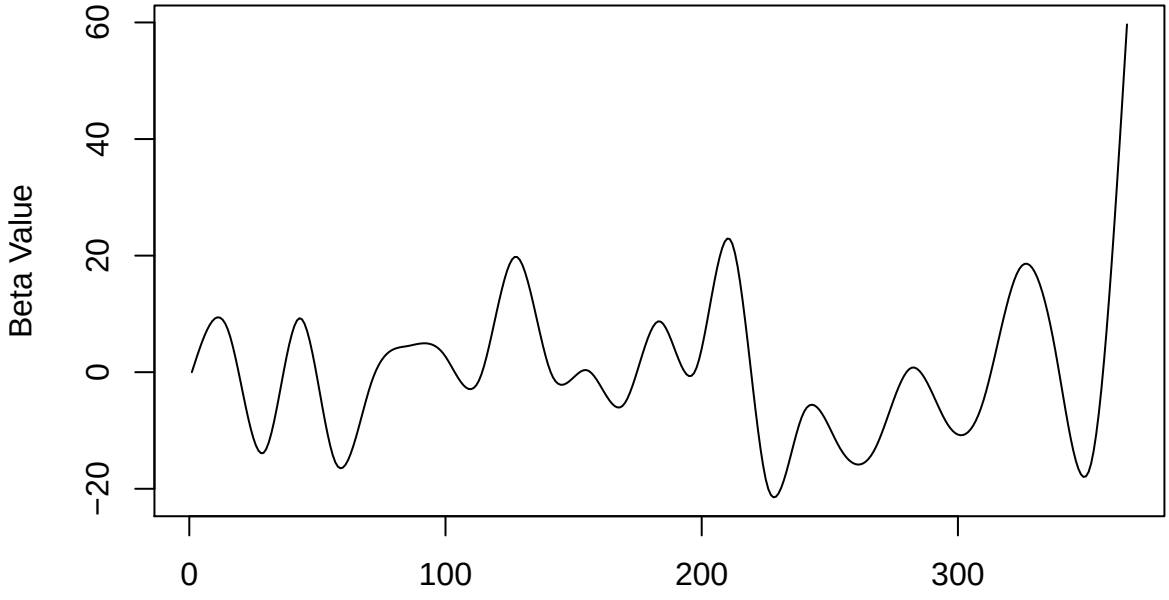
$$Y = X\beta + e,\ e_i\ N(0, \sigma^2)$$

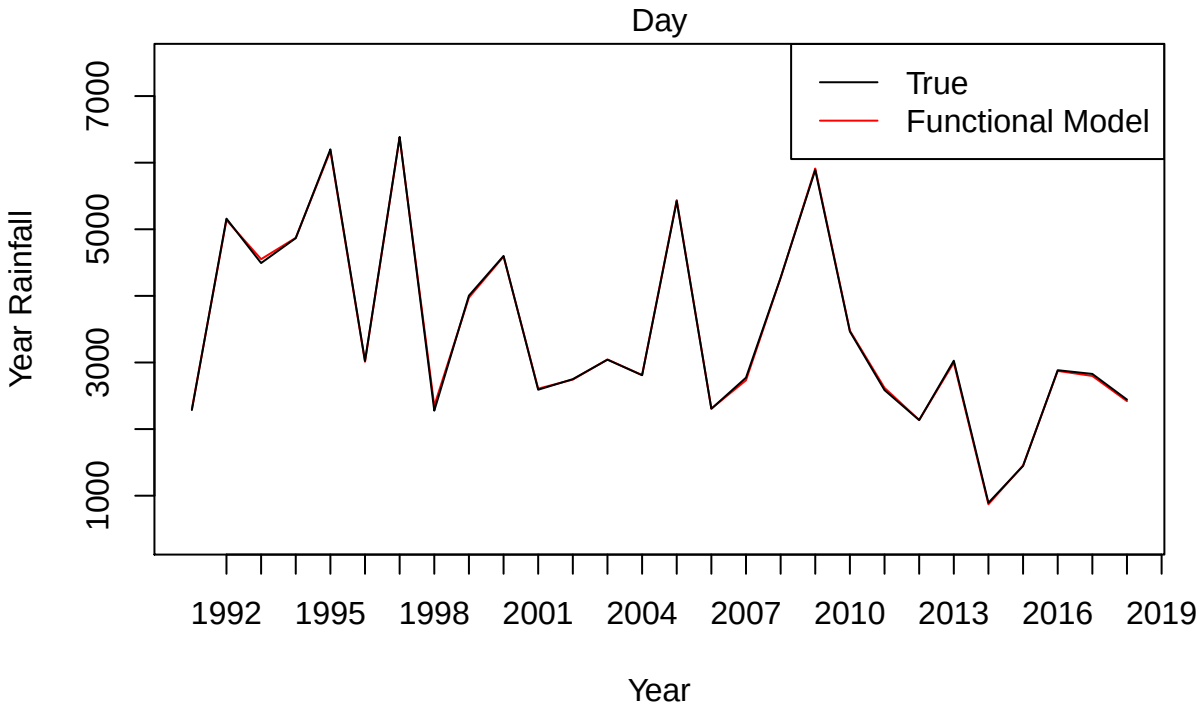
The dimension of X is n by (p+1) for n obseravtions with p predictors. The dimension of β needs to be of dimension (p+1). This leads to a problem with more β values to be evaluated than there are observations(n). This can be resolved by penalising the β coefficients to be on a smooth function, in this case a natural cubic spline. The degree of freedom being 24 resulting in 24 basis functions, there are now fewer coefficients than observations and so the equation can be solved.

To do this β is replaced with $B\alpha$ where B are basis functions. This leaves us with the function

$$Y = XB\alpha + e$$

From this we can regress Y onto XB by finding the α that minimizes the equation, our β values can then be determined by $\beta = B\alpha$. The plot shows the natural cubic spline that the β values were constrained to. The second plot shows the prediced annual rainfall against the true rainfall.





REFERENCES

1. Erni, Birgit, 2020, *Chapter 6: Functional Data Analysis*, lecture notes, Advanced Analytics STA5057W, University of Cape Town, May 2020