# Final STAT547 Spring2014 On the Relation between Temperature and Precipitation from 73 Spanish Weather Station

Yet Nguyen

May 2, 2014

## 1 Introduction

In this project, we will analyze the dataset `aemet` which is available in the package `fda.usc` (Functional Data Analysis and Utilities for Statistical Computing) developed by Manuel Oviedo de la Fuente et al. . The dataset is the series of daily summaries of 73 Spanish weather stations selected for the period 1980-2009. The dataset contains geographic information of each station and the average for the period 1980-2009 of daily temperature, daily precipitation and daily wind speed. The data come originally from Meteorological State Agency of Spain (AEMET) (`http://www.aemet.es/`).

Precipitation is the amount of watter that falls down from clouds. Previous research suggests that the temperature substantially affect precipitation. For instance, according to wikipedia: "during the Last Glacial Maximum of 18,000 years ago, thermal-driven evaporation from the oceans onto continental landmasses was low, causing large areas of extreme desert, including polar deserts (cold but with low rates of precipitation). In contrast, the world's climate was wetter than today near the start of the warm Atlantic Period of 8000 years ago."

In this project, we will look at the relationship between precipitation and temperature under the functional data analysis point of view. In particular, we will investigate the functional structure of those two quantities over daily measures for the period 1980-2009 by principal component analysis. And then we will do a functional linear models in the two forms. The first one is scale response precipitation (calculated as average precipitation over time) vs. temperature curves. The second one is functional response precipitation vs. functional predictor temperature.

## 2 Scatter plot of the two functional datasets

First of all, we can look at the two datasets and their relation which is shown in Figures 1 and 2

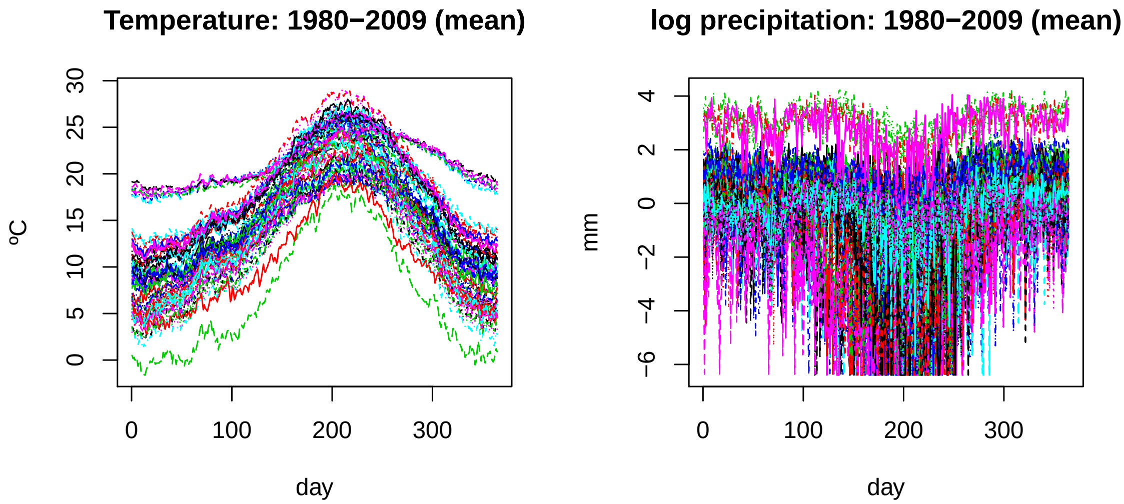


Figure 1: Scatter plot of temperature and log precipitation data

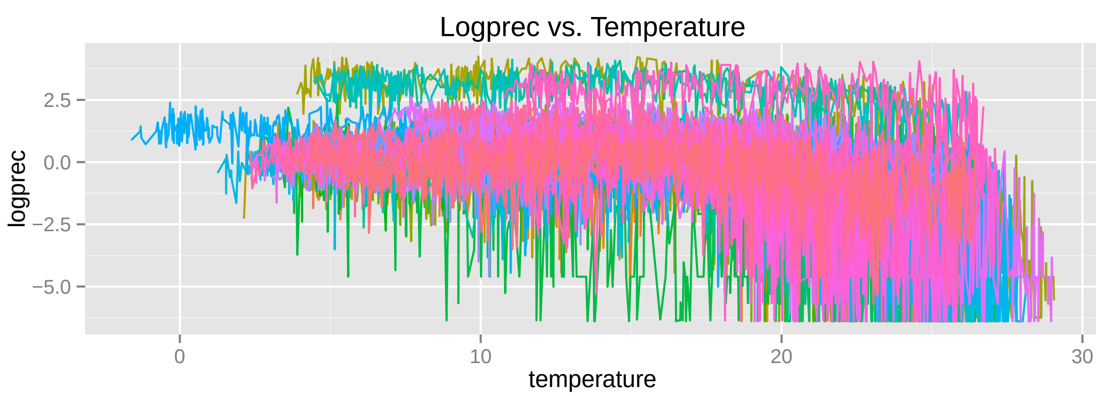


Figure 2: Relationship between temperature and precipitation

# 3 Covariance modelling and Principle Component Analysis for Temperature and Repricipitation

73 curves of each dataset have 365 time points, therefore, we can consider those as *dense* functional data, hence we will use Ramsay and Silverman's Approach to obtain the estimated mean and variance matrix of each data set.

## 3.1 Analysis for Temperature

The estimated mean and covariance is shown in Figure 3.

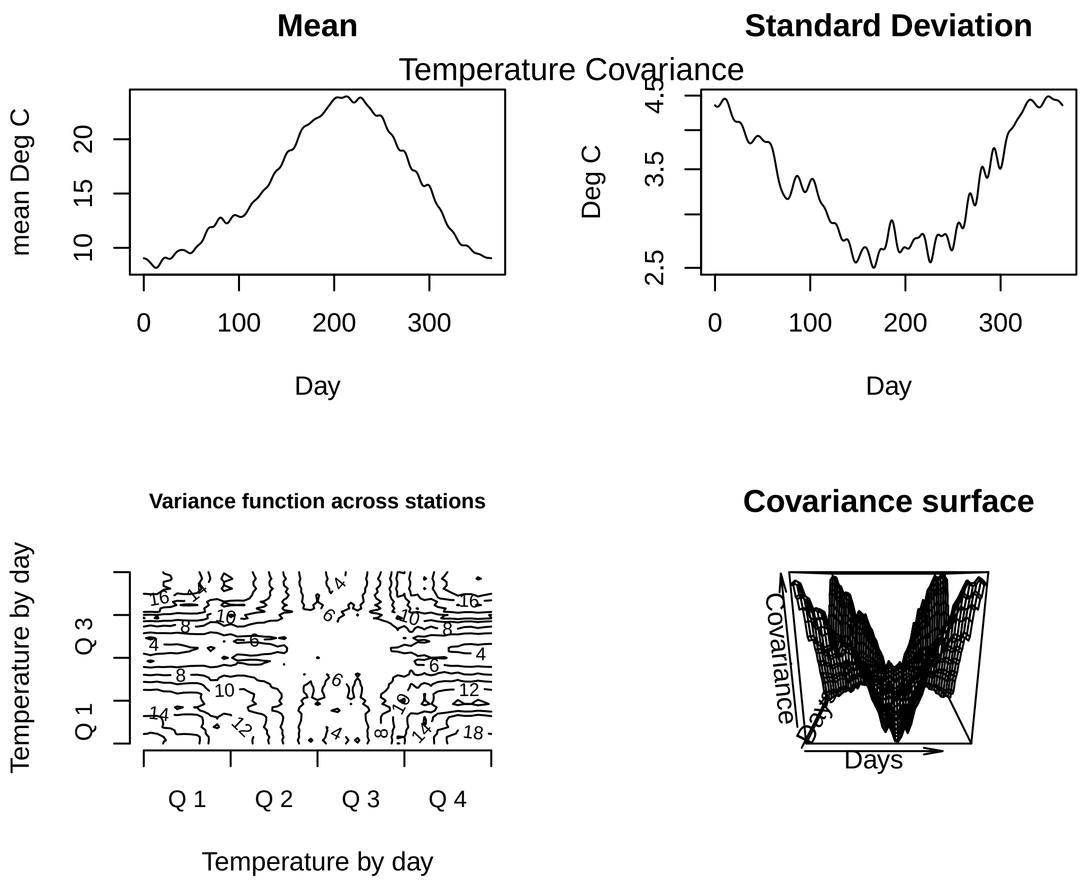


Figure 3: Estimated Mean and Covariance for Temperature data.

The estimated eigenvalues and eigenfunctions of temperature data are shown in Figure 4.

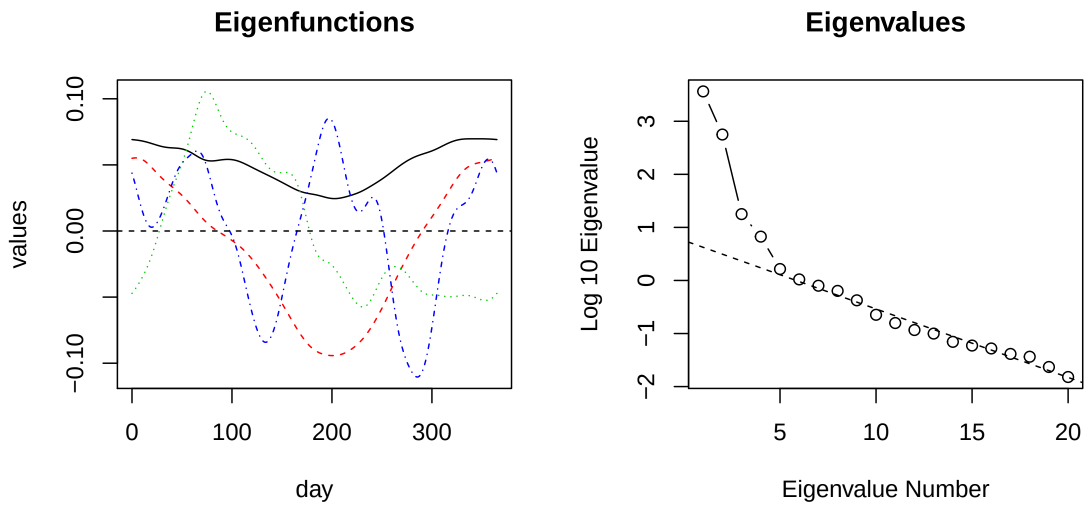


Figure 4: Estimated eigenvalues and eigenfunctions of temperature data.

## 3.2 Analysis for Precipitation

The estimated mean and covariance is shown in Figure 5.

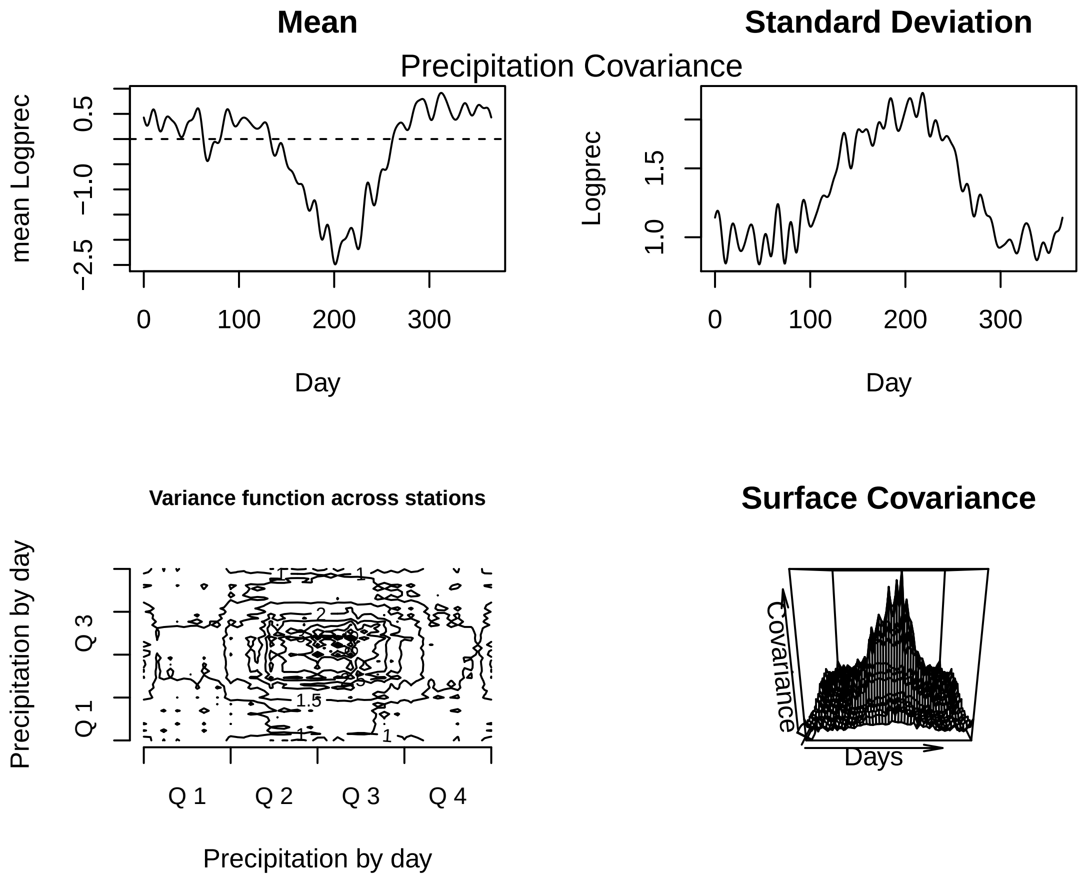


Figure 5: Estimated Mean and Covariance for Precipitation data.

The estimated eigenvalues and eigenfunctions of precipitation data are shown in Figure 6.

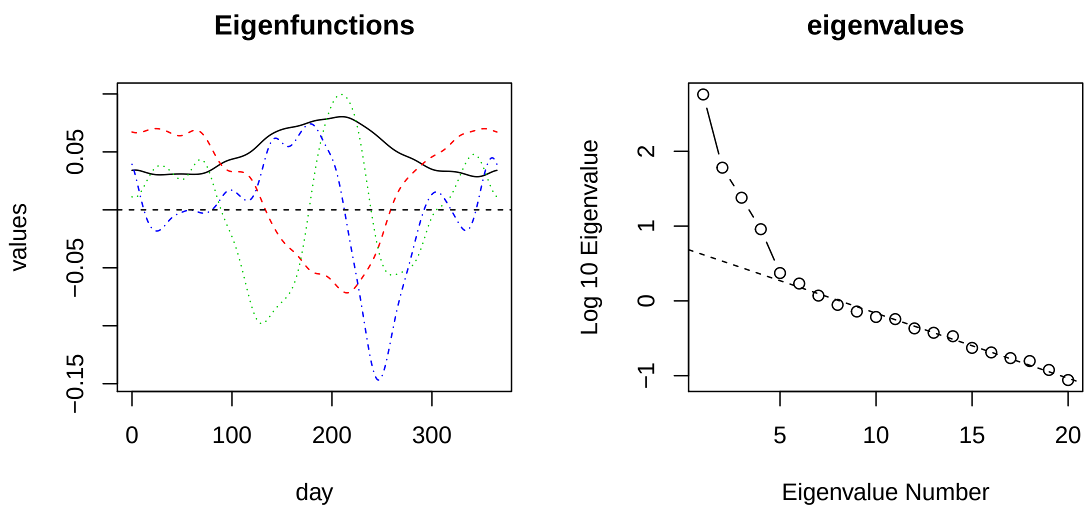


Figure 6: Estimated eigenvalues and eigenfunctions of precipitation data.

Looking at the Figures 3, 4, 5, and 6, we can see that the behavior of *temperature* and *precipitation* are differently in the opposite direction, i.e., somehow "negative" correlation, both in the average mean curve and covariance surface.These conclusions do support the hypotheses that there is negative correlation between temperature and precipitation. In the next sections, we will investigate the functional regression relation between these two functional datasets.

# 4 Functional Linear Model with Functional Covariate as Temperature and Scalar Response as Log of Total Precipitation:Ramsay and Silverman's Approach and Principle Component Analysis Approach

Figure 7 shows the scatter plot of log of total precipitation.

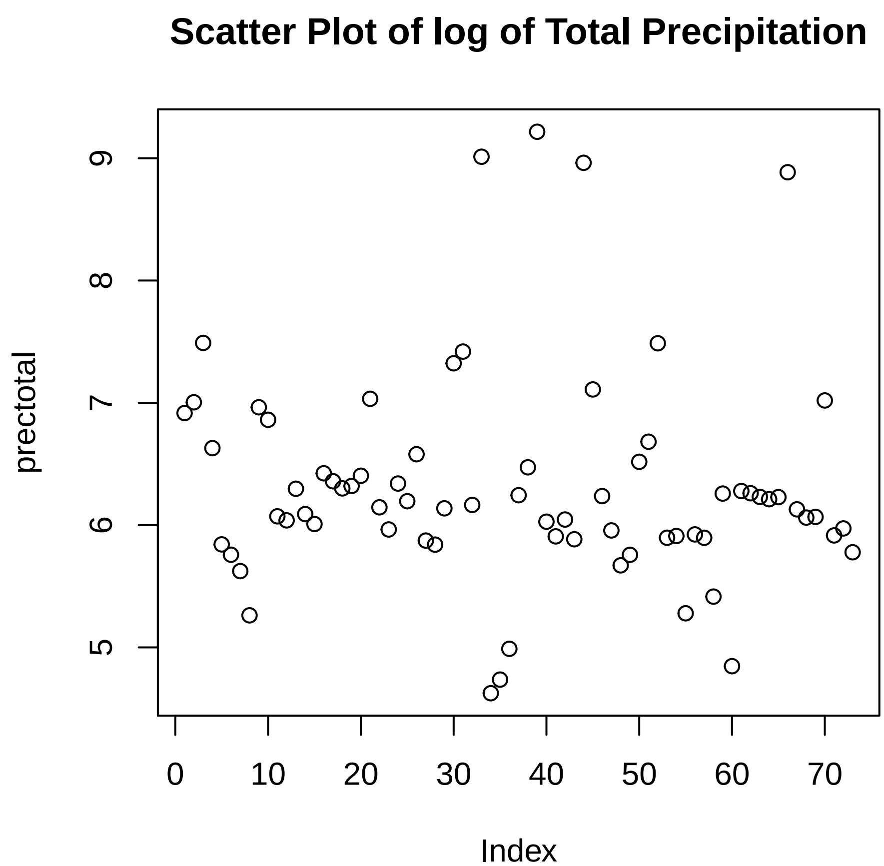


Figure 7: Scatter plot of Log of Total Precipitation.

In this section, we will fit a functional linear model where scalar response is the logarithm of total of precipitation, and functional covariate is temperature curves using Ramsay and Silverman's approach. In particular, we choose Bspline basis for the functional covariate, and without penalty. The generalized cross validation is used to obtain optimal number of basis functions. The result of mean square error for the prediction is 2.081e-15.

Moreover, we also use Principle Component Analysis to fit the linear model,the selection is performed by cross-validation to choose the best principle components.

```
## [1] 2.081e-15
## [1] 1.219e-17
```

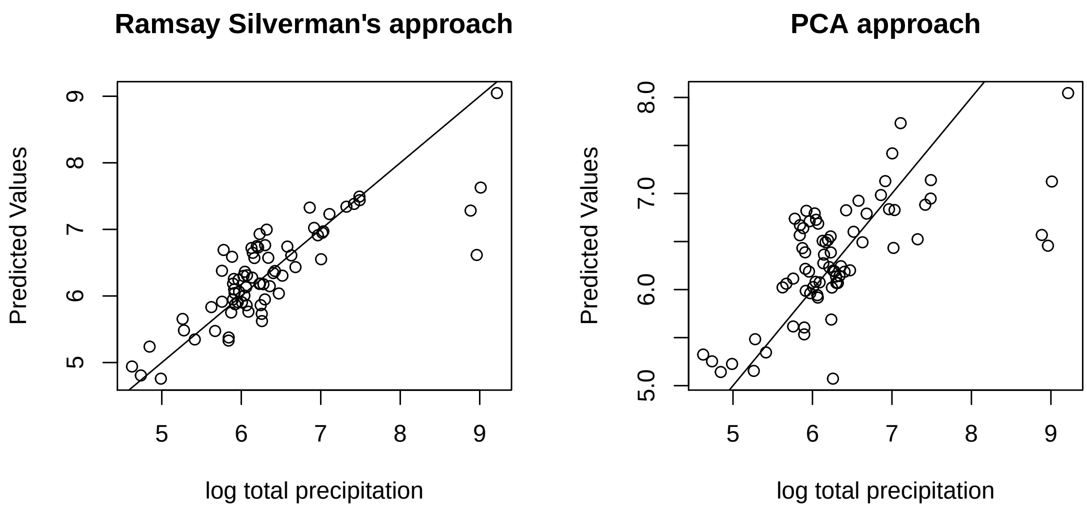


Figure 8: Fitted Values by Ramsay Silverman s Approach and PCA Approach.