Predicting Droughts in the Amazon Basin based on Global Sea Surface Temperatures

Dario Lepke

Contents

Introduction

Placeholder

Chapter 1

Related work

Placeholder

Chapter 2

EDA

Placeholder

2.1 EDA precipitation

2.2 Glyph plots

2.3 EDA SST

Chapter 3

Correlation analysis

Placeholder

3.1 Short Recap

3.2 Correlation of Sea Surface Temperature and Precipitation

3.2.1 Original Data

3.2.1.1 Timelag 0

3.2.1.2 Timelag 3

3.2.1.3 Timelag 6

3.2.1.4 Timelag 12

3.2.2 Deseasonalised Data

3.2.2.1 Timelag 0

3.2.2.2 Timelag 3

3.2.2.3 Timelag 6

3.2.2.4 Timelag 12

3.3 Summary

3.3.1 Original Data

3.3.2 Deseasonalised Data

Chapter 4

Clustering

In this chapter we will first summarize the main ideas of clustering and then apply it to the precipitation data. If not indicated otherwise the information is taken from Elements of Statistical Learning.

4.1 Main Idea Clustering

We can describe an object by a set of measurements or its similarity to other objects. Using this similarity we can put a collection of objects into subgroups or clusters. The objects in the subgroups should then be more similar to one another than to objects of different subgroups. This means inside the clusters we aim for homogeneity and for observations of different clusters for heterogeneity. With the clustering analysis applied to the precipitation data we want to study if there are distinct groups (regions) apparent in the CAB. So that if we later apply the regression models we predict the precipitation for each group and not for the whole region.

To explore the grouping in the data we need a measure of (dis)similarity. This measure is central and depends on subject matter considerations. We construct the dissimilarities based on the measurements taken for each month. We interpret this as a multivariate analysis where, each month is one variable. So given the area in the CAB (resolution 5°x5°), we have 612 cells and 432 months, resulting in a 612×432 data matrix. we want to cluster cells into homogen groups.

4.2 Clustering Methods

4.2.1 K-means

4.2.2 Kmeans characteristics

4.2.3 K-medoids

4.2.3.1 K-medoids characteristics

4.2.4 PCA

4.2.5 Gap statistic

4.3 Analyse clustering results

Chapter 5

LASSO Regression

Placeholder

5.1 The LASSO

5.2 Optimization

5.3 TODO here

Chapter 6

lasso center

Placeholder

6.1 LASSO model

6.2 Error plots

6.3 Coefficient plots

6.4 Inspect predictions from each fold

6.5 Inspect predictions from best CV-lambda

6.6 Summary

Chapter 7

lasso stand

Placeholder

7.1 LASSO model

7.2 Error plots

7.3 Coefficient plots

7.4 Inspect predictions from each fold

7.5 Inspect predictions from best CV-lambda

7.6 Summary

Chapter 8

lasso center

Placeholder

8.1 LASSO model

8.2 Error plots

8.3 Coefficient plots

8.4 Inspect predictions from each fold

8.5 Inspect predictions from best CV-lambda

8.6 Summary

Chapter 9

lasso center

Placeholder

9.1 LASSO model

9.2 Error plots

9.3 Coefficient plots

9.4 Inspect predictions from each fold

9.5 Inspect predictions from best CV-lambda

9.6 Summary

Chapter 10

The fused lasso

Placeholder

10.1 General

10.1.1 Implementation