Master's Thesis

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

Department of Statistics
Ludwig-Maximilians-Universität München

Dario Lepke

Munich, August 9th, 2022

Submitted in partial fulfillment of the requirements for the degree of M. Sc.
Supervised by Dr. Fabian Scheipl (LMU) and Dr. Niklas Boers (PIK)

Introduction

With future climate change droughts in the Amazon forest may become more frequent and/or severe. Droughts can turn Amazon regions from rain forest into savanna, leading to high amounts of carbon released into the atmosphere. Therefore, predicting future droughts and understanding the underlying mechanisms is of great interest. Ciemer et al. (2020), established an early warning indicator for droughts in the central Amazon basin (CAB), based on tropical Atlantic sea surface temperatures (SSTs). Inspired by their work, the aim of this thesis is to build up on this work and improve its predictive power by using different statistical methods. Here we seek to build a model that is able to predict monthly precipitation based on the sea surface temperatures. Also we want to identify those sea regions that are most important for doing so, making interpretability a point of interest, too. Firstly we will analyze the data descriptively to explore patterns and spatial dependencies. This includes a cluster analysis of the precipitation data in the central Amazon basin. Following we will compare two different regression approaches and their capability to predict precipitation as well as their interpretability of the SST regions selected by them. The first model is the lasso as proposed by Tibshirani (1996). Comparing different model specifications we will carry on the findings from the lasso and fit a (sparse) fused lasso on the data (Tibshirani et al. (2005)). Both models will be evaluated using a 5-fold forward selection, a model evaluation technique that takes into account the time dependencies present in the data at hand. We conclude with a summary of the findings in this work and give an overview of strengths and limitations of the approaches used together with ideas for future research.

This thesis was written and supervised in cooperation with Dr. Niklas Boers from the Potsdam Institute for Climate Impact Research (Climate Impact Research (PIK) e. V. (2022)) and Dr. Fabian Scheipl (LMU)

Related work

Placeholder

EDA

Placeholder

EDA precipitation

Glyph plots

EDA SST

Correlation analysis

Placeholder

Short Recap

Correlation of Sea Surface Temperature and Precipitation

Original Data

Timelag 0

Timelag 3

Timelag 6

Timelag 12

Deseasonalised Data

Timelag 0

Timelag 3

Timelag 6

Timelag 12

Summary

Original Data

Deseasonalised Data

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.

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.

Clustering Methods

k-means

k-means characteristics

K-medoids

K-medoids characteristics

PCA

Gap statistic

Analyse clustering results

LASSO Regression

Placeholder

Introduction

Implementation

TODO here

Results

lasso

standardized lasso

deseas lasso

diff1 lasso

The fused lasso

Placeholder

Introduction

Implementation

Ciemer, Catrin, Lars Rehm, Juergen Kurths, Reik V Donner, Ricarda Winkelmann, and Niklas Boers. 2020. "An Early-Warning Indicator for Amazon Droughts Exclusively Based on Tropical Atlantic Sea Surface Temperatures." *Environmental Research Letters* 15 (9): 094087.

Climate Impact Research (PIK) e. V., Potsdam Institute for. 2022. "Potsdam Insitute for Climate Impact Research." 2022. https://www.pik-potsdam.de/en.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–88.

Tibshirani, Robert, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. 2005. "Sparsity and Smoothness via the Fused Lasso." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (1): 91–108.