Recap MA Thesis Lepke

Dario Lepke

30.11.2021

Introduction

Main objective

- ▶ Predict drought in Central Amazonas Basin (CAB) based on global sea surface temperatures (SST)
- ▶ Motivation: Droughts in Amazonas strongly influence regional ecosystem and lead to high biomass carbon impact
- ▶ Reference work: "An early warning indicator for Amazon droughts exclusively based on tropical Atlantic SST" (Ciemer et al. 2020)

Summary reference paper

Ciemer et al, 2020

- Monthly data from 1981 to 2016, 35 years
- SST: Compute monthly anomalies w.r.t long term mean
- Drought: Precipitation -> Drought (3-SPI), then averaged over Central Amazon Basin
- Compute correlations for SST and drought, over whole period of time
- Identify 4 highly correlated regions (unweighted networks, 10% strongest correlations)

Correlation Map

Positive correlations

Negative correlations
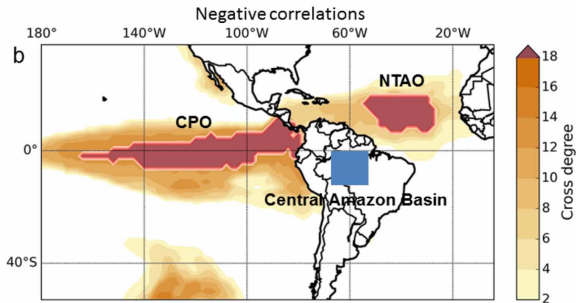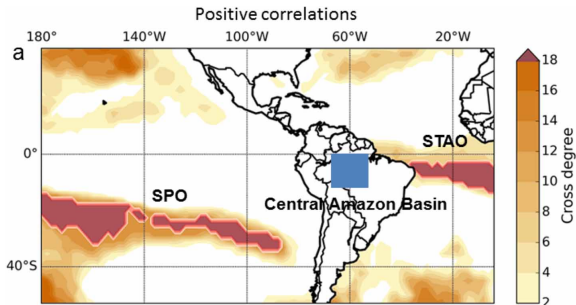
Methods

- Use significant correlations as weights in further analysis
- For each region create series of networks
- Each network based on 24 month of data, sliding window
- Result is time series of Average Cross Correlation (ACC)
- Information from whole data set used for the sliding window approach
- Two atlantic regions become more interesting (NTAO and STAO)
- Investigate ACC between NTAO and STAO

ACC of NTAO, STAO and drought index
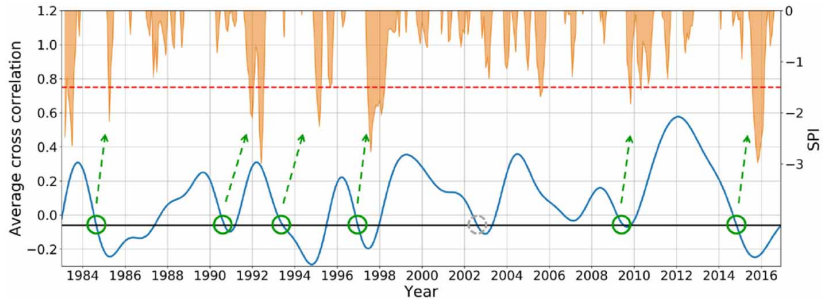
Methods summarised

- SST vs drought
- Identify 4 oceanic regions
- Explore relationship of 4 regions with drought over time
- Further study relationship over time between 2 regions in atlantic (NTAO and STAO only)
- Study relationship of ACC (NTAO, STAO) and drought

Methods summarised

- Weights from whole time period get used as weights for shorter time periods in sliding window
- Predictive outcome is dichotomous (drought yes/no)

Master Thesis

Planned work

- Run explorative Analysis
- Fit a predictive model to the data
- LASSO regression, FUSED LASSO, variable preselection...
- Apply statistical methods for model evaluation/ validation (Cross Validation for Time Series)
- Use different time lags to study development over time

Summary work done so far

- Using CDO for data handling
- Applying STL algorithm
- Correlation Analysis before and after STL
- Cross Validation for Time Series
- First LASSO Models fitted

Corr Plots

- Showing corr-analysis document

Model Pipeline so far

- Deseasonalise complete data via STL (!)
- Average precipitation in CAB
- Discard final 5 years for validation (!)
- Run Blocked Cross Validation (5-folds)
- Fit LASSO model in each run

CV fold plots

- showing CV fold plots

Planned Improvements

- ▶ Further EDA (clustering analysis, only use significantly correlated regions)
- ▶ First discard final 5 years then run model pipeline in each fold:
  - ▶ Precipitation to drought
  - ▶ Deseasonalise with STL
  - ▶ Average drought in CAB
  - ▶ Fit and test LASSO model
- ▶ Validate on last 5 years
- ▶ Different models, timelags, CV measures

Open

- ▶ Drought instead of precipitation?
- ▶ Influence of different CV approaches
- ▶ Other ML models