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

Dario Lepke

Contents

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). In my thesis I would like to build on this work and improve the predictive power by using different statistical methods. Meaning, we seek to build a model that is able to predict droughts (resp. rainfall) based on preceding sea temperatures, desirably with as much lead time as possible. Also we want to identify those sea regions that are most important for doing so, making interpretability a point of interest, too. A first model could be a cross-validated (generalised) LASSO approach trying to identify the most important oceanic regions and the respective time-scales.

The thesis will be done in cooperation with Dr. Niklas Boers from the Postdam Institute for Climate Impact Research (Climate Impact Research (PIK) e. V. (2021)).

Chapter 1

Related work

As already mentioned the paper by Ciemer et al. (2020), created an early warning indicator for Amazon droughts. They did so using a complex network approach. They used two datasets, one for the Sea Surface Temperatures (Smith et al. (2008)) and one for the precipitation (Funk et al. (2015)) with monthly data for the time period of 1981 until 2016. The data can be downloaded for example in netcdf format and manipulated conveniently with Climate Data Operators (CDO, Schulzweida (2019)). CDO in turn can be used with wrappers for R and Python. The data is organized on a longitude/latitude grid.

They identify 4 oceanic regions that correlate the most with rain in the amazon basin, using a coupled network approach. Figure 1 shows the cross degree towards rainfall in the central Amazon basin (CAB, blue box), for positive and negative correlations. Darker shades indicate a larger cross degree, hence a larger number of links and correlations with rainfall at more grid points in the CBA.

The correlations are measured using a spearman rank-order correlation coefficient

$$\rho = 1 - \frac{6 \sum \Delta_{R_i}^2}{n(n^2 - 1)}. \tag{1.1}$$

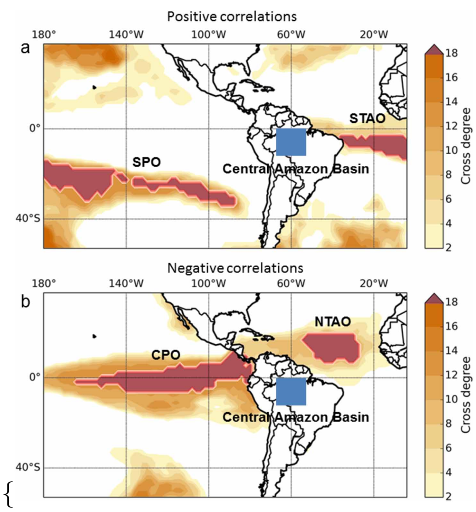
Where Δ_{R_i} denotes the difference between the ranks of observations of both variables at the same time i and n is the number of observations. An Adjacency Matrix describes the resulting network, where the threshold p_{th} was chosen so that only 10% of the strongest correlations are represented as links in the network

$$A_{ij} = \begin{cases} 0 \; p_{ij} < p_{th} \\ 1 \; p_{ij} \geq p_{th} \end{cases} . \tag{1.2}$$

The cross degree then gives the strength of correlation between a specific grid point i of network V_l (oceanic grid point) and another (sub)network V_m (all grid points j in central amazon basin)

$$k_i^{lm} = \sum_{j \in V_m} A_{ij}, i \in V_l \ . \tag{1.3}$$

\begin{figure}

{ 

}

\caption{"Cross degree between sea surface temperature and continental rainfall anomalies. For each sea surface temperature grid cell of the Atlantic and Pacific Ocean, the cross degree towards rainfall in the Central Amazon Basin (blue box) is shown, for a positive correlations and b negative correlations. Darker shading indicates a larger cross degree, implying a larger number of links, and thus significant correlations with rainfall at more grid points in the Central Amazon Basin. Red areas outline coherent oceanic regions with a the 20% highest cross degrees for positive correlations, found in the Southern Pacific Ocean (SPO) and Southern Tropical Atlantic Ocean (STAO), and b the 20% highest cross degrees for negative correlations, found in the Central Pacific Ocean (CPO) and Northern Tropical Atlantic Ocean (NTAO)" (Ciemer et al. 2020)} \end{figure}

They further explore the relationship by constructing (weighted) networks for sliding windows of 24 months between the Central Amazon Basin and each of the ocean regions. For each month except for the first two years, an individual network is computed based on the data of the previous 24 months. Then they take the average of the cross correlations for each of the networks which gives a new time series of average cross correlation (ACC) values. Each ACC summarizes the connectivity of one region with the CAB for the last 24 months.

They find that NTAO and STAO give the strongest signal, hence they apply the same sliding window coupled network approach between the ocean regions NTAO and STAO. Before they computed networks between ocean and continental regions, now it is computed between these two Atlantic regions, NTAO and STAO.
The resulting time series and its comparison to the drought index time series is shown in figure 1.1 below. They find that using a ACC threshold for a drought (SPI below -1.5), lets them forecast 6 of the 7 droughts in the observation period, missing the 2005 drought, while also giving one false alarm in 2002.

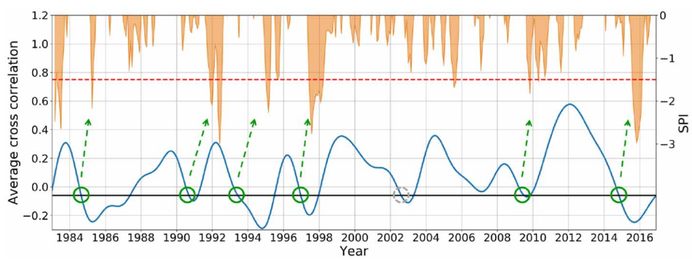


Figure 1.1: "Early-warning signal for droughts in the central Amazon basin. We compare the time evolution of the average cross correlation of the Northern Tropical Atlantic Ocean (NTAO) and Southern Tropical Atlantic Ocean (STAO), given by the blue curve, with the standardized precipitation index (SPI, orange) of the central Amazon basin. Orange dips indicate a negative SPI with a threshold for severely dry periods (SPI -1, dotted red line). We expect a drought event within the following one and a half years whenever the average cross correlation between NTAO and STAO SST anomalies falls below an empirically found threshold of -0.06. Green circles indicate a matching forecast based on the Atlantic SST correlation structure, with one false alarm in 2002 indicated by a grey circle, where the threshold is crossed but no drought took place in the direct aftermath (see Discussion). The temporal evolution of the average cross correlation shown here is smoothed using a Chebyshev type-I low-pass filter with a cutoff at 24 months" [@ciemer2020early].

The work by Ciemer et al. (2020) shows potential forecasting capabilities and limitations. While they are able to predict 5 out of 6 drought events, they also give one false negative and one false positive result. Their work uses a complex network approach that is applied stepwise (first two unweighted networks, then two weighted networks and in the end a dichotomous threshold decision rule). For the thesis we would like to create a more general predictive model that can learn the relationship between the SSTs and rainfall in the CAB. As already mentioned a first step can be a LASSO model. First findings show that, the classic LASSO only chooses single points in the ocean as predictors, though. But our motivation is to discover predictive regions and not only single separated points. Therefore in a next step we want to make use of a generalized form of the LASSO that also takes into account that chosen predictors should be close to each other. This model is the so called Fused LASSO.

For the models we also need a form of evaluation. Classic Cross Validation assumes independence of the observations. In our setting this is clearly violated due to the time dependency of the data. We will explore different possibilities to use an adjusted form of Cross Validation that takes this characteristic into account.

Depending on how well the relationship between SST and rain can be established, we can take this a step further and use it as a so called “Emergent Constraint” (EC). Since different climate models give different answers about future climate there is a need to narrow this “spread”, which can be done by ECs. To do so, we need a plausible relationship between a Variable X and Y (here: SST and drought).

According to how well the relationship is represented in a climate model we assign “credibility” to a climate model’s future projections (here: projections of future droughts in the Amazon rain forest). In summary this can be used to reduce uncertainty in the ensemble of climate models’ future projections, f.e by using ML techniques as done by Schlund et al. (2020).

Chapter 2

EDA

2.1 EDA precipitation

In this section we want study the time series of precipitation in the Central Amazon Basin. The CHIRPS data set contains the **precipitation data**, created from in-situ and satellite measurements (Funk et al. (2015)). It can be downloaded for example from here [https://www.chc.ucsb.edu/data/chirps]. It contains observations **from 1981 to 2021**. Data comes on a **high resolution of 0.05 grid**.

Time frame: 1980 to 2016

2.1.1 Overview

Below the area of the Central Amazon Basin that is object of our study.

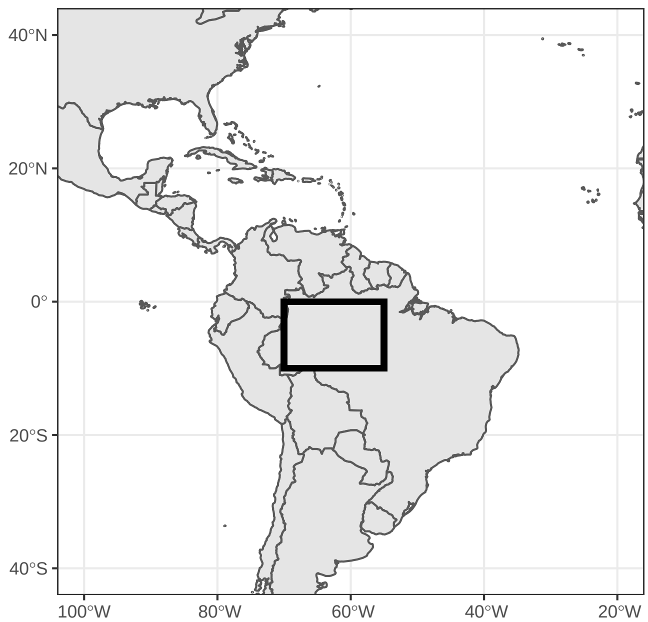
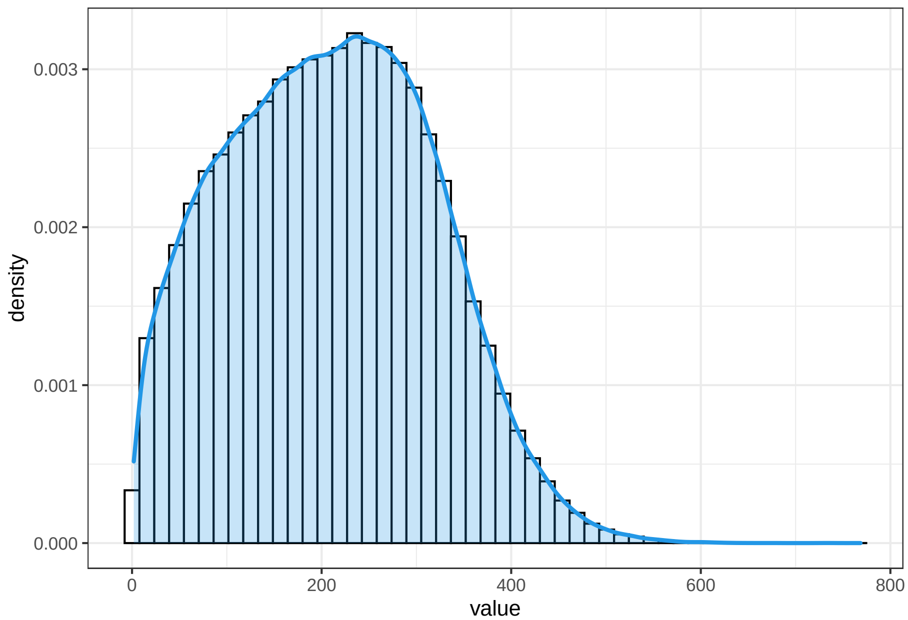


Figure 2.1: Localisation of the area under study. The central amazon basin (CAB) spanning across 0,-10 latitude and -70,-55 longitude

2.1.2 Precipitation values raw

Firstly we inspect the precipitation values in general



Its form is a unimodal, right-skewed distribution. The values range from `r[1]` tom `r[2]`.

2.1.3 Mean at each location

As we can see most locations have a mean precipitation of around 200 mm/month, over the whole time series. Regionally in the "upper left" corner of the Amazon Basin, mean precipitation is higher or equal to the mean. The reference point for "higher" is the mean of the location means. This region seems to be more or less spatially consistent. The rest of the region with lower mean precipitation has also some small areas where precipitation is again a little bit higher. For example in the upper right corner and on the bottom, right of the middle.

2.1.4 SD at each location

For the standard deviation we also see regional patterns. These patterns overlap with the regions of the mean but their magnitude is flipped. Meaning, in the upper left where we observe larger mean values we generally observe lower standard deviation and in the lower and upper right corners, higher standard deviations.

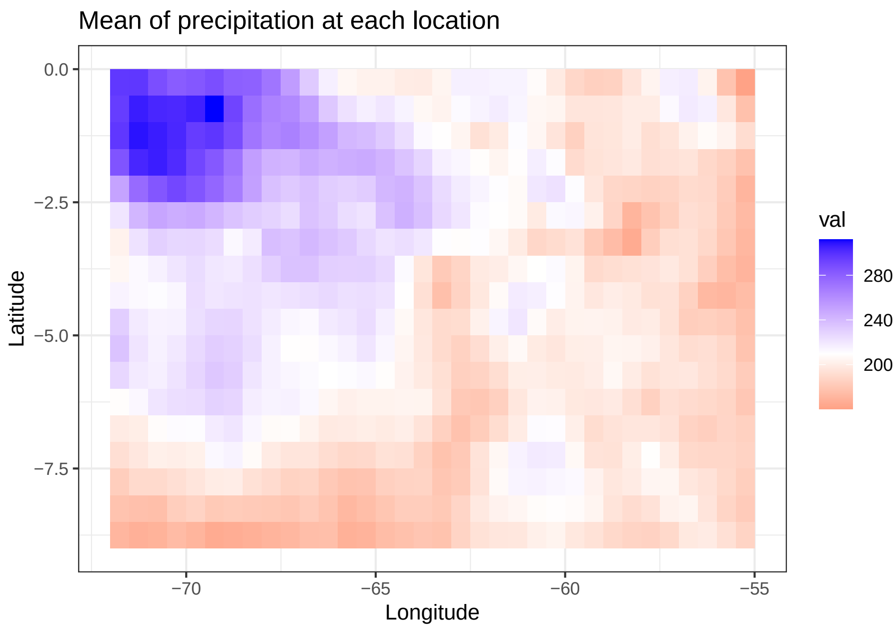


Figure 2.2: Precipitation mean at each location. The mean was computed over the whole time period

Question: There are obviously regional differences in magnitude of mean and standard deviations. Should we therefore NOT normalise the time series, prior to clustering? Mean and SD contain information and the variables are all measured on the same scale. (Consideration: seasonality might play a role, meaning that the values of different months come a different distribution)

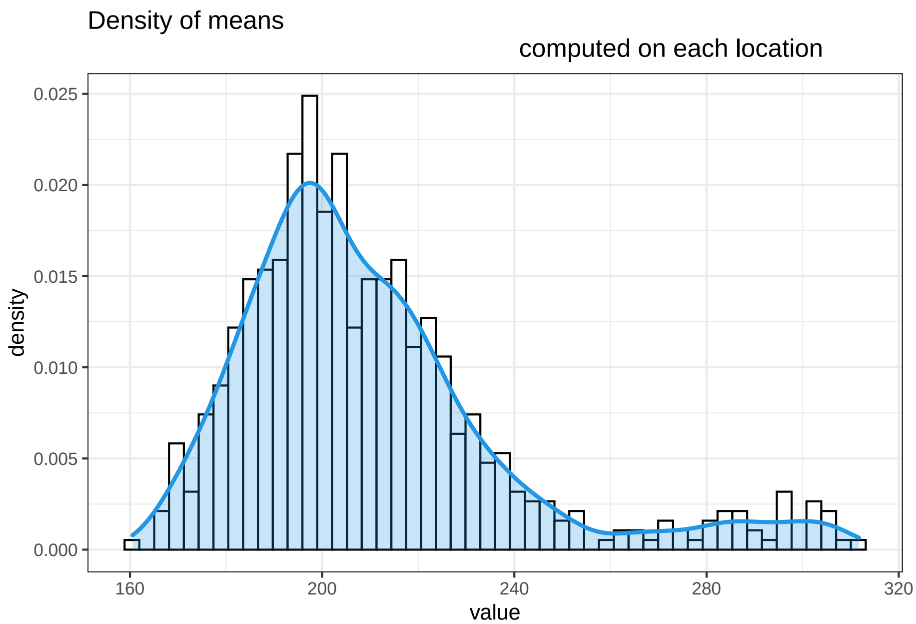


Figure 2.3: Density of means, means were computed for each location over the whole time period

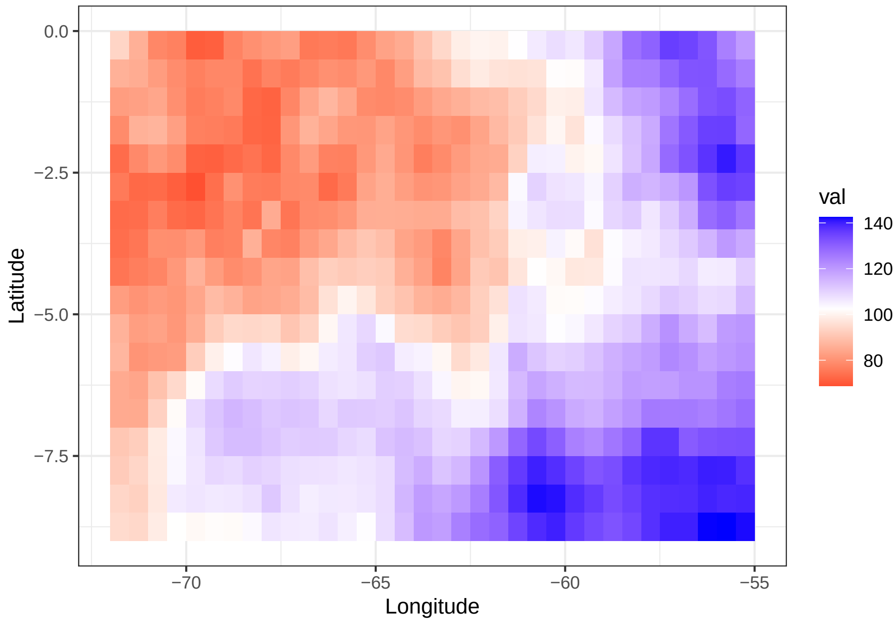
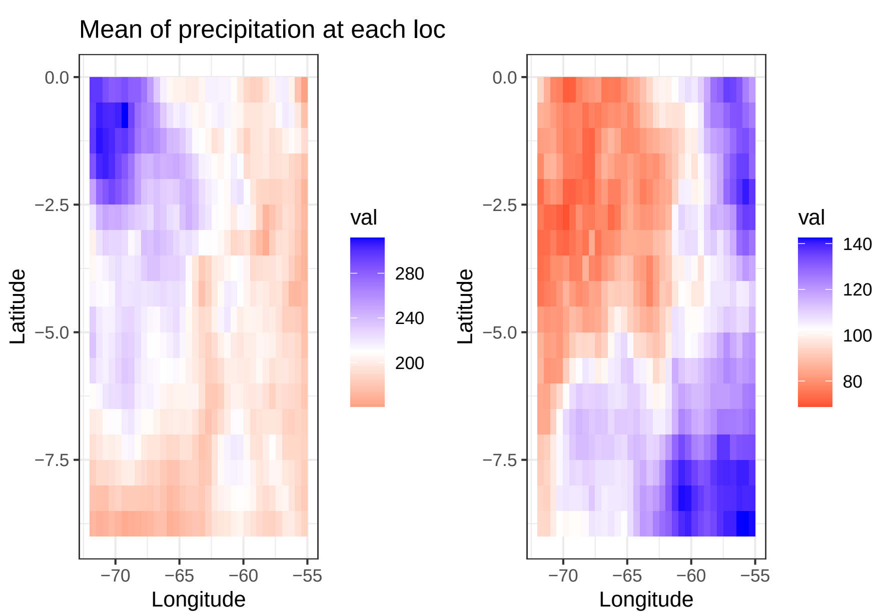
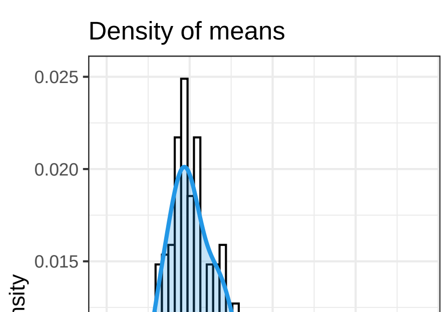
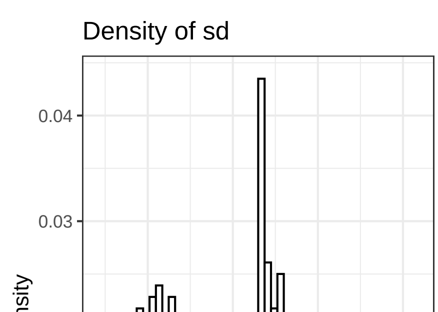


Figure 2.4: Precipitation standard deviation at each location. The standard deviation was computed over the whole time period

2.1.5 Mean and SD at each location







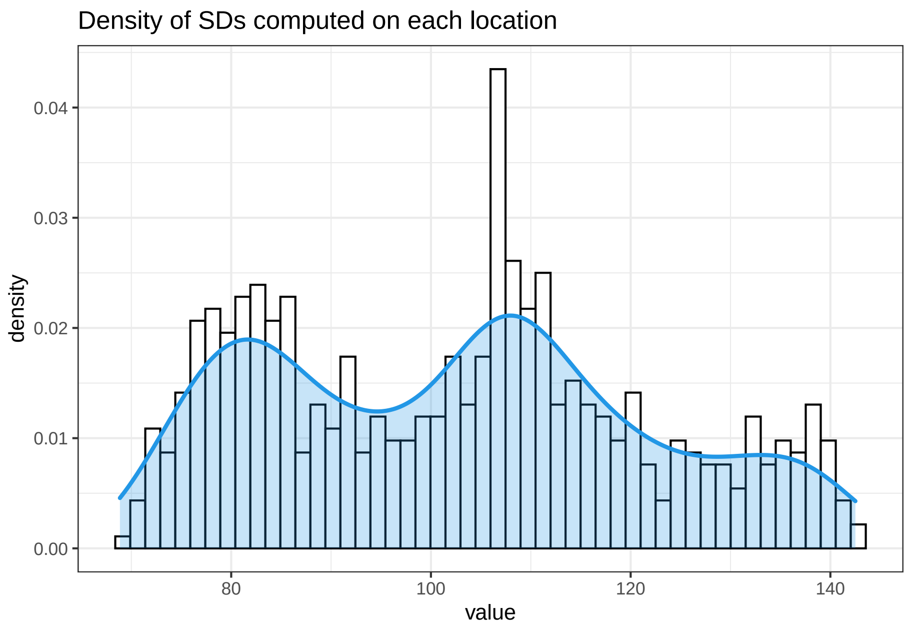
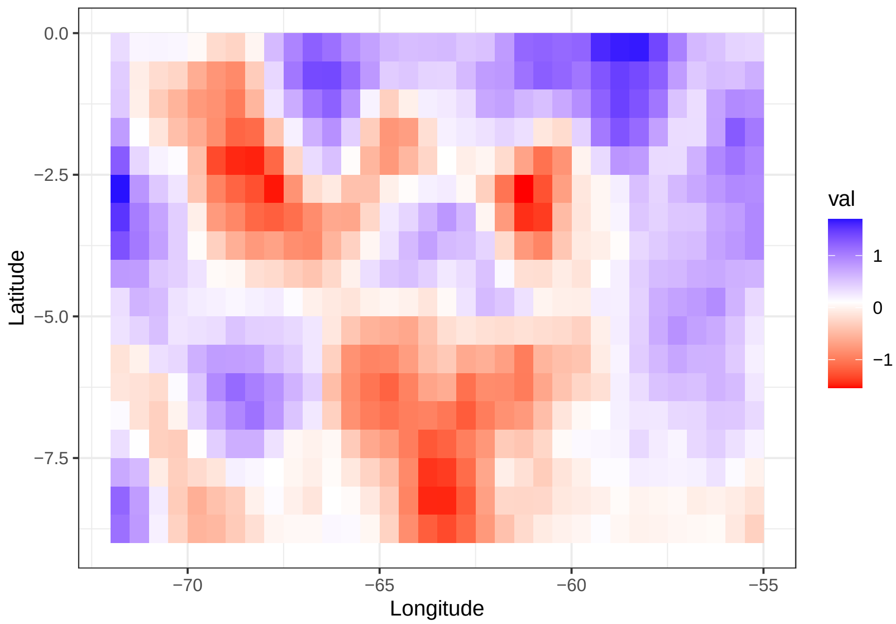


Figure 2.5: Density of standard deviations, standard deviations were computed for each location over the whole time period

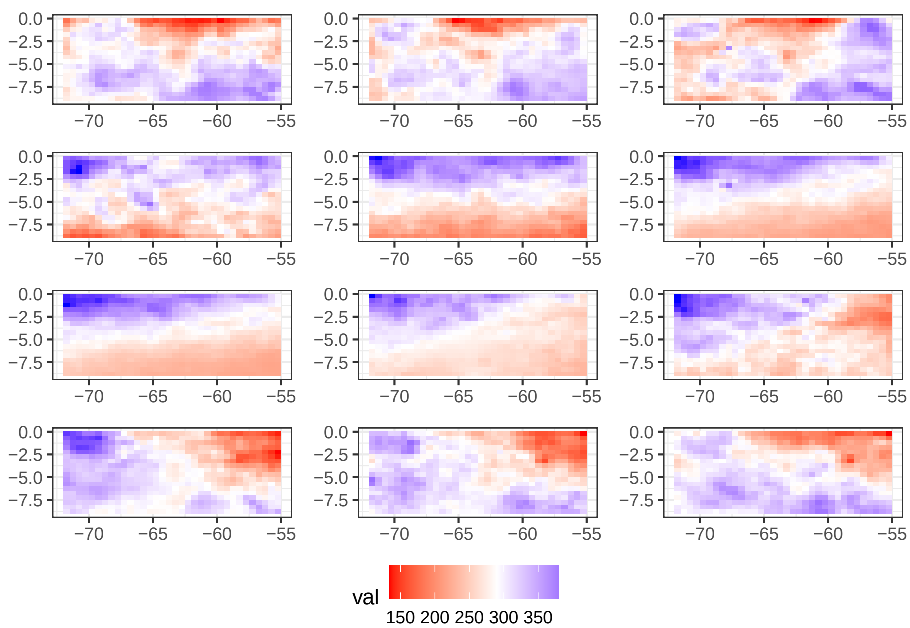
2.1.6 Trend at each location



Many regions dont have a trend and trends in general seem to be quite small. We can identify some regions with similar up or downward trends. We also computed plots for the deseasonalised data but it looked almost the same

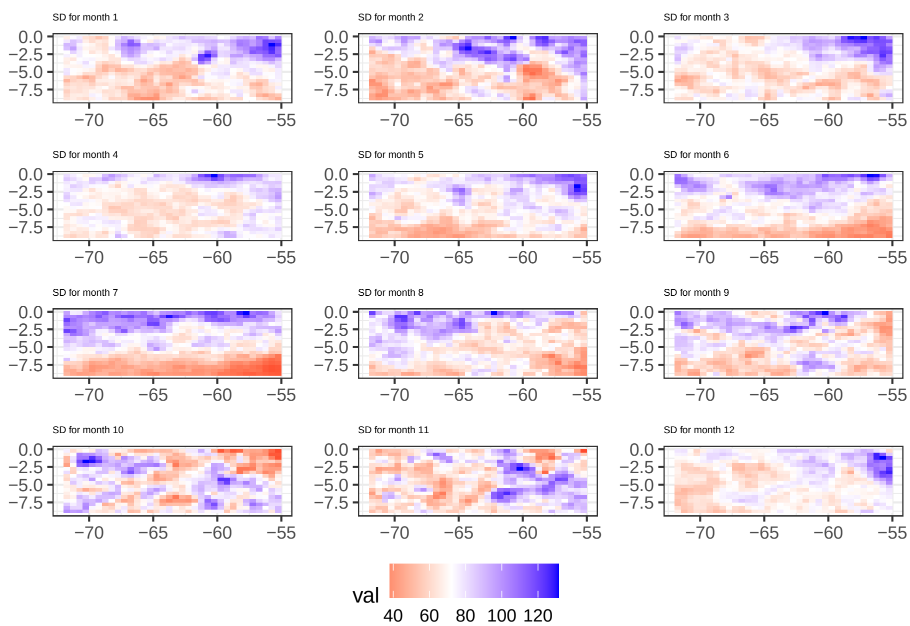
Plot means and sd for each month respectively

2.1.7 Means per month TEST



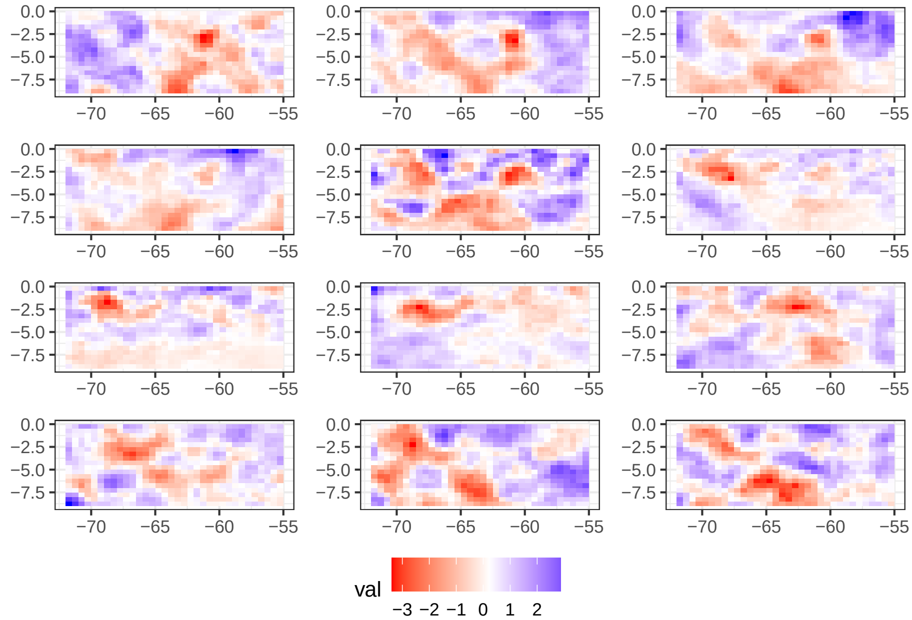
We see spatial patterns of the mean evolving over time. For example: From May until August there is a spatial separation in two parts that dissolves in september. As expected there is a large seasonal component regarding the means.

2.1.8 SD per month TEST



For the standard deviation we see as well large differences in values during different months of the year.

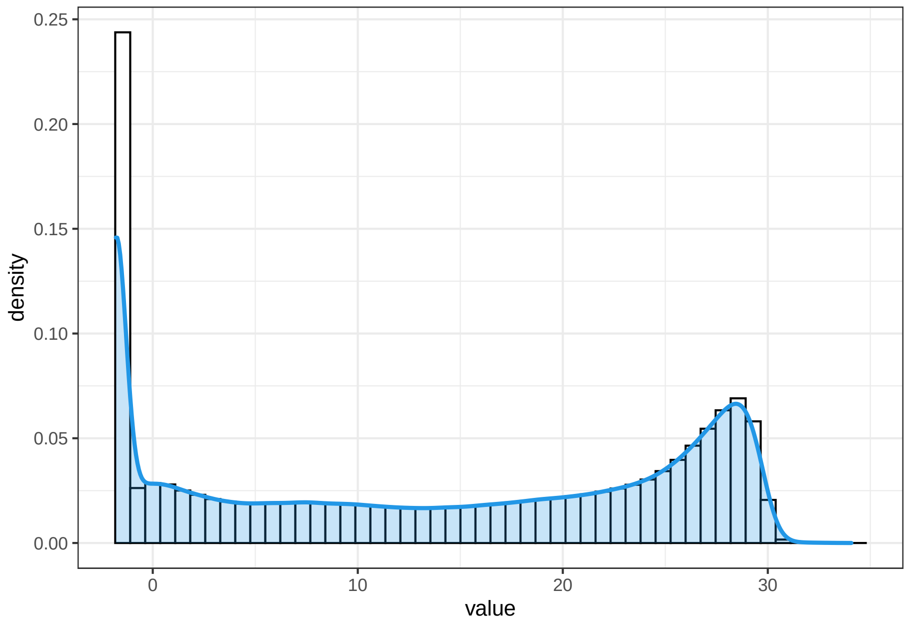
2.1.9 Trend per month TEST



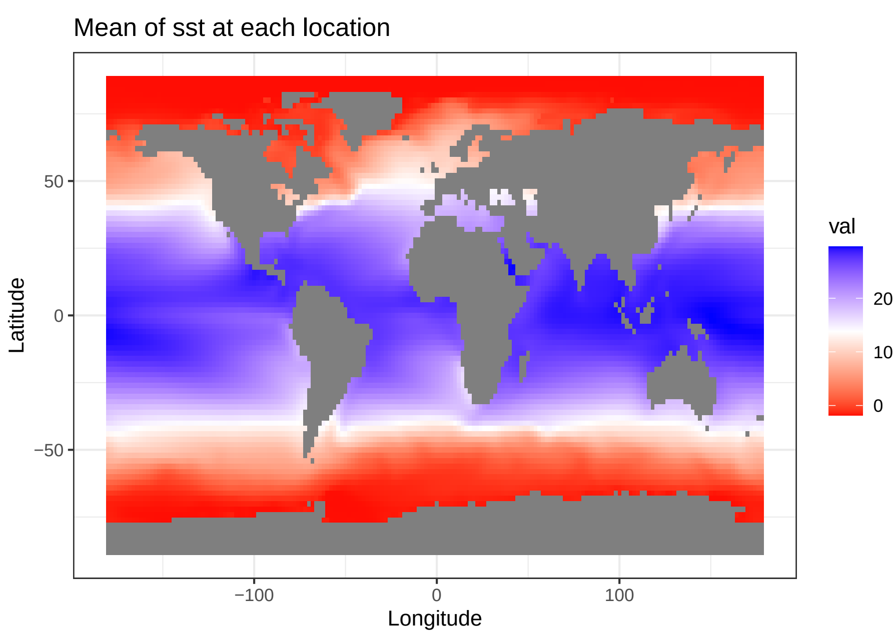
2.2 EDA SST

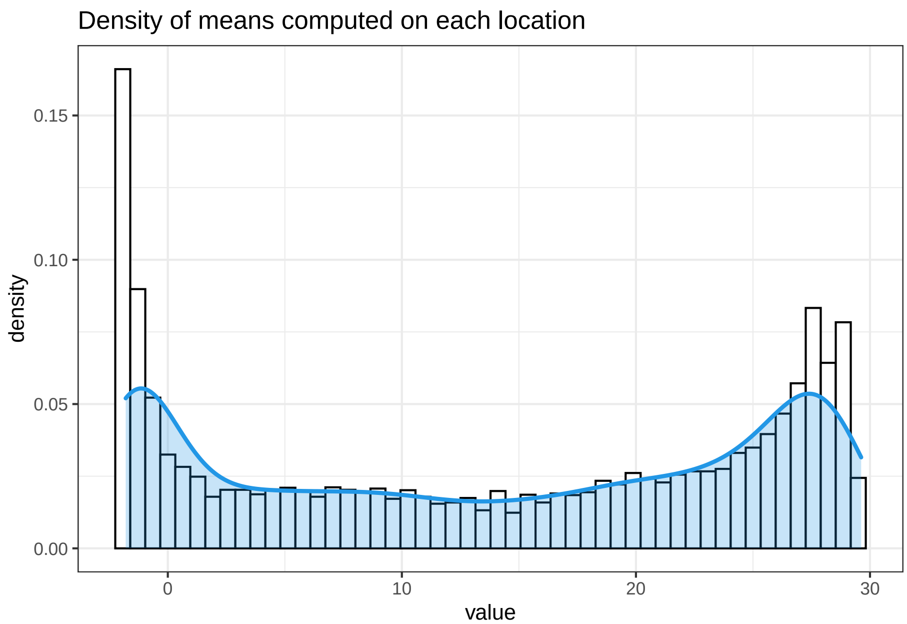
We explore the **sea surface temperature data** set, used in the paper by Ciemer et al (Ciemer et al. (2020)). ERSST (Extended Reconstructed Sea Surface Temperature, Huang et al. (2017)) is a reanalysis from observed data given in the International Comprehensive Ocean-Athmosphere Data Set (ICOADS). Which contains observations **from 1800/01 until 2016/12**, made by ships and buoys for example. The data comes on a **2x2 degree grid**, where data was missing **interpolation techniques were used**. See paper for reference. the file contains **two variables** that are measured across different dimensions. The two variables contain the **sea surface temperatures and the respective SST anomalies** (with respect to the 1971-2000 monthly climatology).

2.2.1 SST values raw



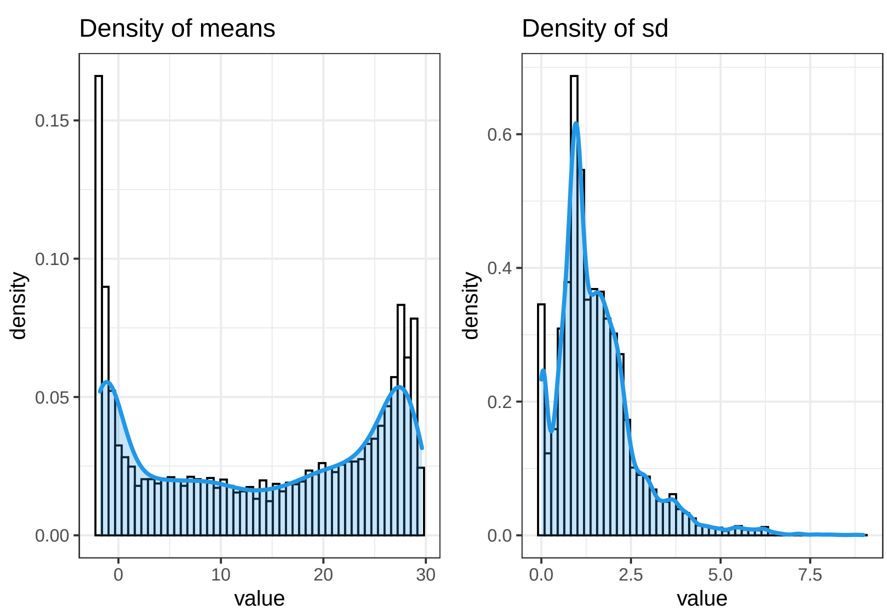
2.2.2 Mean at each location





2.2.3 SD at each location

2.2.4 Mean and SD at each location



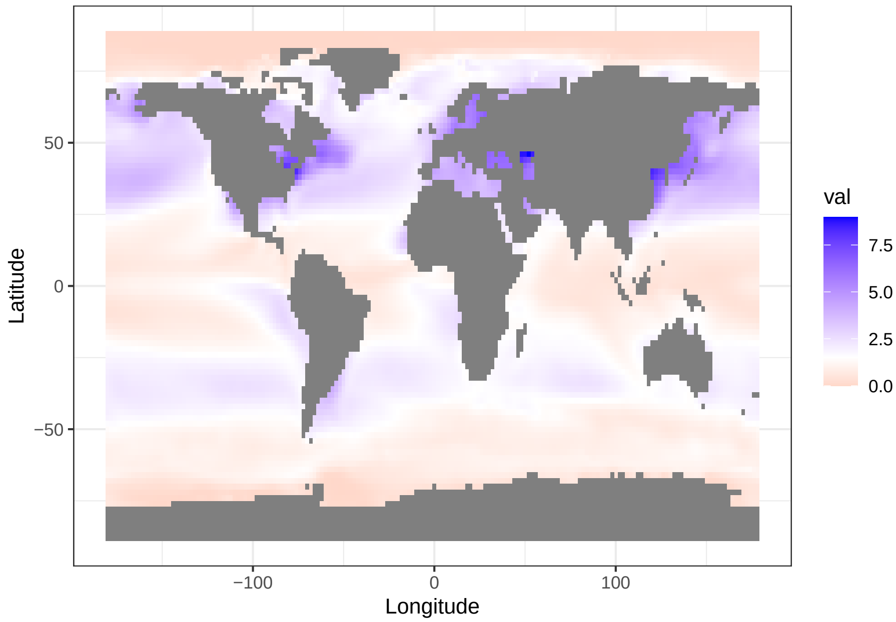


Figure 2.6: SST standard deviation at each location. The standard deviation was computed over the whole time period

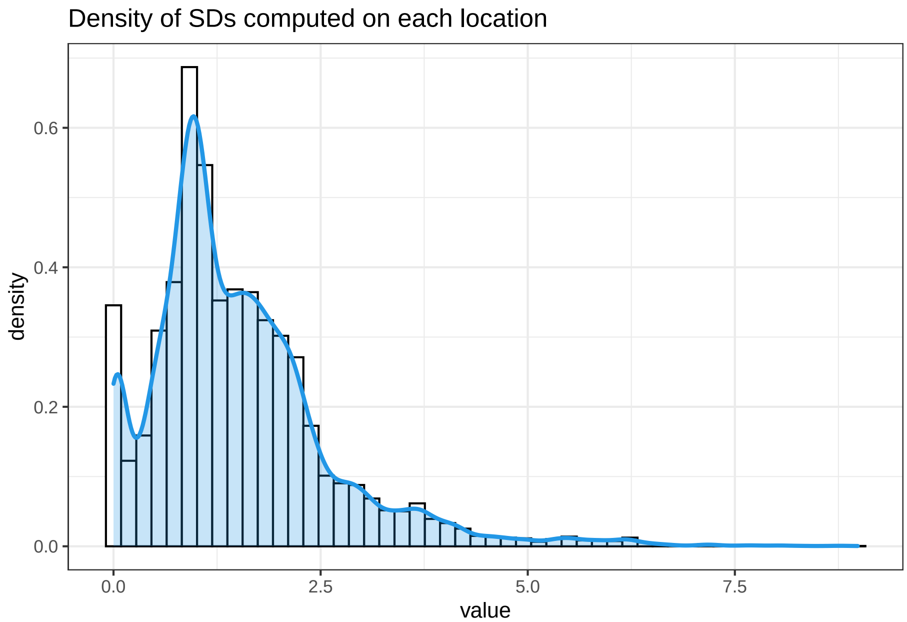


Figure 2.7: Density of standard deviations, standard deviations were computed for each location over the whole time period

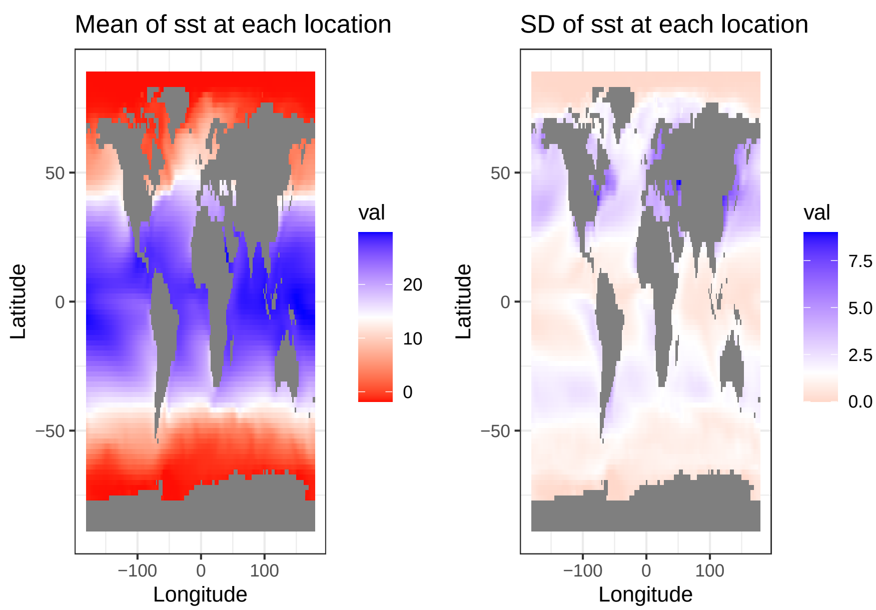
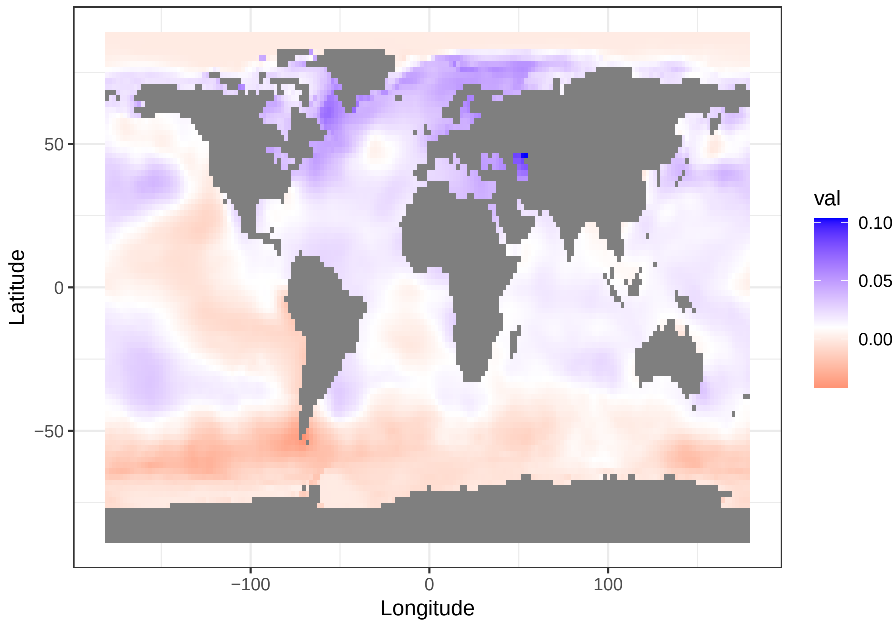
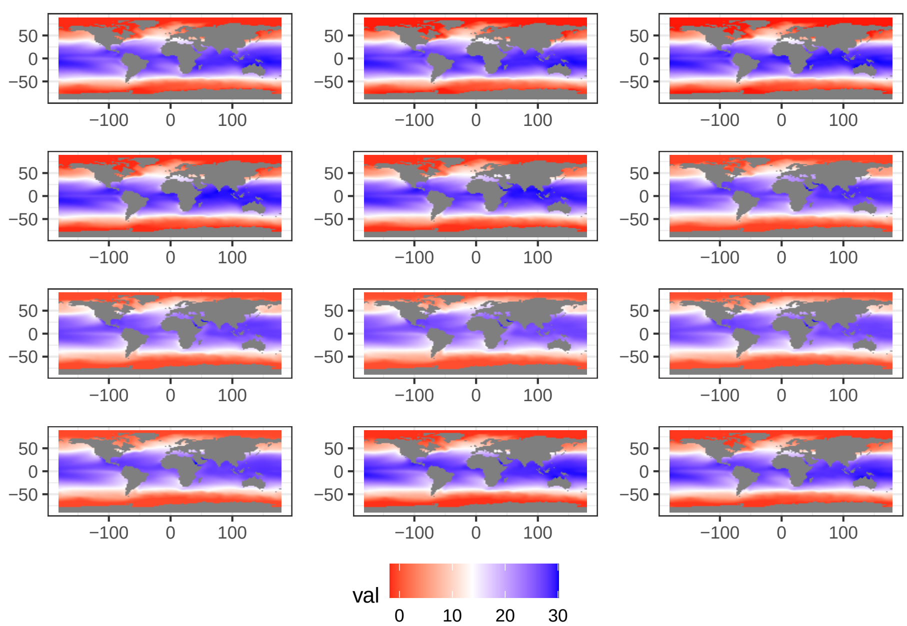


Figure 2.8: Mean and SD on the spatial map.

2.2.5 Trend at each location

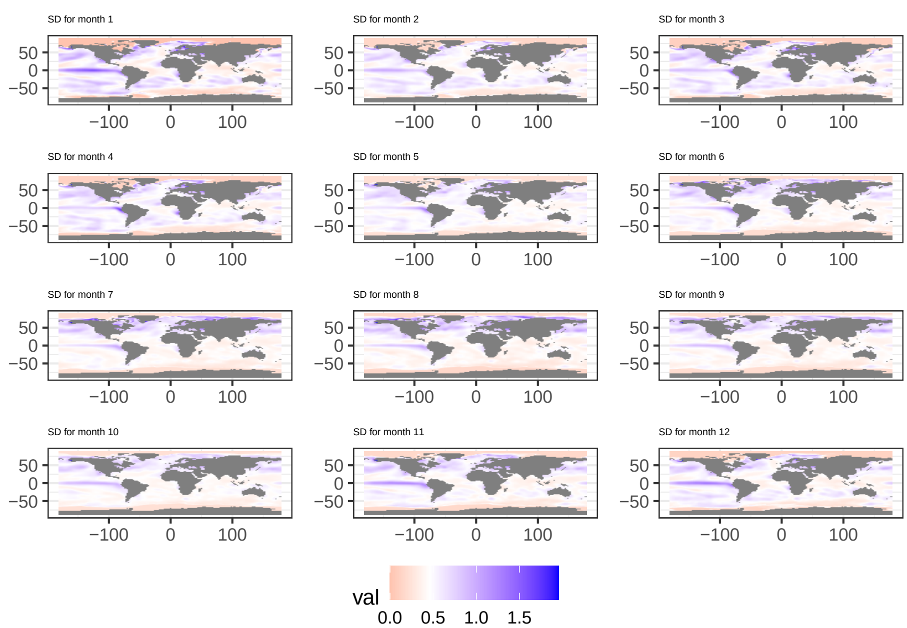


Means per month TEST



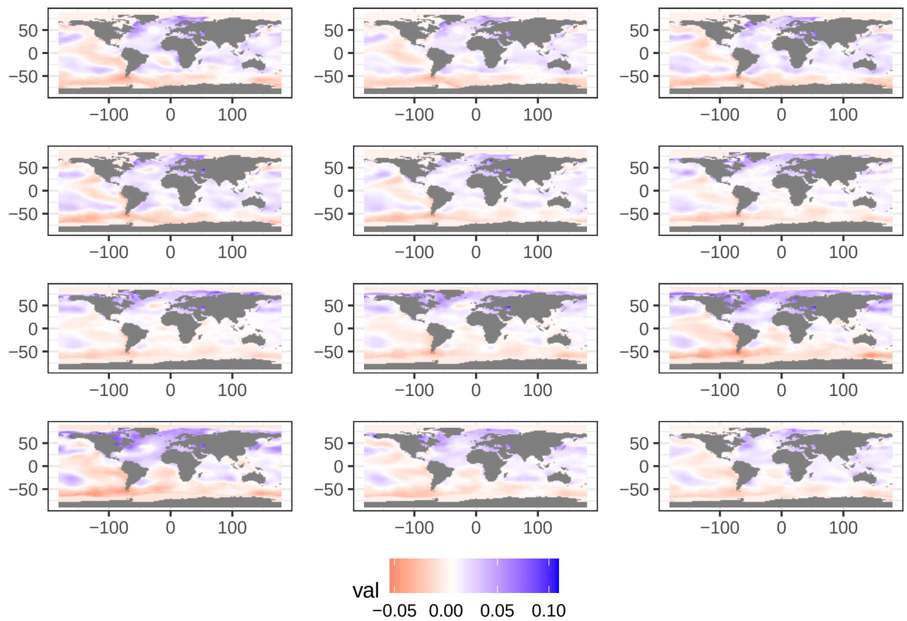
We see spatial patterns of the mean evolving over time. For example: From May until August there is a spatial separation in two parts that dissolves in september. As expected there is a large seasonal component regarding the means.

2.2.6 SD per month TEST



For the standard deviation we see as well large differences in values during different months of the year.

2.2.7 Trend per month TEST



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. 2021. "Potsdam Insitute for Climate Impact Research." 2021. https://www.pik-potsdam.de/en.

Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, et al. 2015. "The Climate Hazards Infrared Precipitation with Stations-a New Environmental Record for Monitoring Extremes." *Scientific Data* 2 (1): 1–21.

Huang, Boyin, Peter W Thorne, Viva F Banzon, Tim Boyer, Gennady Chepurin, Jay H Lawrimore, Matthew J Menne, et al. 2017. "NOAA Extended Reconstructed Sea Surface Temperature (ERSST), Version 5." *NOAA National Centers for Environmental Information* 30: 8179–8205.

Schlund, Manuel, Veronika Eyring, Gustau Camps-Valls, Pierre Friedlingstein, Pierre Gentine, and Markus Reichstein. 2020. "Constraining Uncertainty in Projected Gross Primary Production with Machine Learning." *Journal of Geophysical Research: Biogeosciences* 125 (11): e2019JG005619.

Schulzweida, Uwe. 2019. "CDO User Guide (Version 1.9. 6)." *Max Planck Institute for Meteorology: Hamburg, Germany.*

Smith, Thomas M, Richard W Reynolds, Thomas C Peterson, and Jay Lawrimore. 2008. "Improvements to NOAA's Historical Merged

Land–Ocean Surface Temperature Analysis (1880–2006).” *Journal of Climate* 21 (10): 2283–96.