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

Dario Lepke
Dr. Fabian Scheipl (Ludwig-Maximilians-University Munich)
Dr. Niklas Boers (Potsdam Institute for Climate Impact Research)

September 08, 2022

Fused lasso without sub-graphs, gamma 0.1

Fused lasso without sub-graphs, gamma 0.05

Fused evaluation (maybe explain this when showing resutls)

Introduction

Motivation

- The Amazon basin is a key hotspot of biodiversity, carbon storage and moisture recycling
- Hydrological extremes affect ecosystem and populations tremendously
- Droughts in the Amazon rainforest can have severe biomass carbon impact
- Severe Amazon drought in 2010 had total biomass carbon impact of 2.2 PgC, affected area $3 miokm^2$

Related work

- Ciemer et al. (2020) established an early warning indicator for water deficits in the central Amazon basin (CAB)

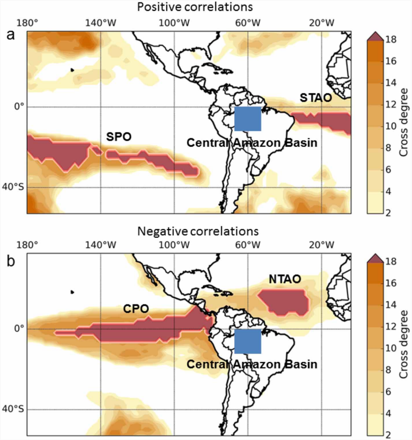


Figure 1: Cross degree between sea surface temperature and continental rainfall anomalies. For each grid cell of sea surface temperature in the

Early warning signal

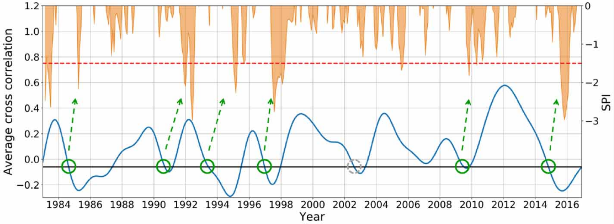


Figure 2: Early-warning signal for droughts in the central Amazon basin. We compare the time evolution of the average cross-correlation of the Northern Tropical Atlantic Ocean (NTAO) and Southern Tropical Atlantic Ocean (STAO), given by the blue curve, with the standardized precipitation index (SPI, 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

Our Approach

- Inspect spatial and temporal characteristics in raw data
- Directly predict rain from SST
- Use lasso and fused lasso
- model evaluation with cross validation for time series

Explorative analysis

The Data

- Rain data from CHIRPS (Climate Hazards Group InfraRed Precipitation woth Station data)
- CHIRPS contains in-situ and satellite data
- SST data from ERSST (Extended Reconstructed Sea Surface Temperature)
- ERSST is reanalysis of observation data (made by ships and buoys for example), missing data filled by interpolation techniques
- These are the same data sets as in Ciemer et al. (2020)

Explorative Analysis Rain

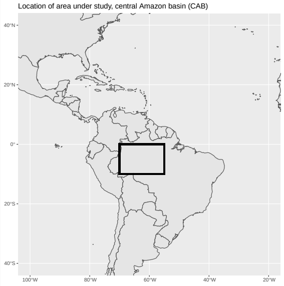


Figure 3: Location of the area under study. The central amazon basin (CAB) spanning across 0,-10 latitude and -70,-55 longitude

Precipitation, Mean and SD

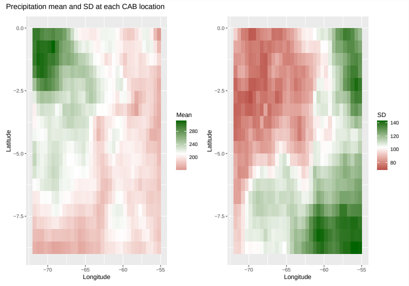


Figure 4: Mean and standard deviation at each location. The standard deviation was computed over the whole time period. The white line on the scale at the side of the plots indicates the mean of the respective quantity

Precipitation Glyph Plots

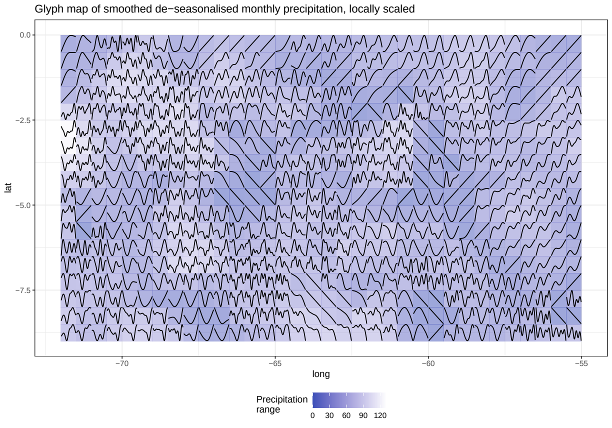


Figure 5: Glyph map of de-seasonalised and smoothed precipitation. The time series are scaled locally, ranges are not the same in all cells. The different ranges are given in color shades, where lighter shading indicates a larger range and darker shades smaller ranges.

Explorative analysis SST

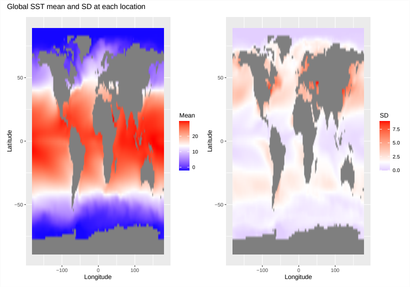


Figure 6: Mean and SD on the global map.The color scales show the mean for the shown variable as white.

Correlation analysis

Correlation analysis

- Here we get an overview over the general correlation structure of the complete data
- We show correlation data for the original as well as the seasonally adjusted data
- The seasonal component was removed by using the stl algorithm that separates the time series into

$$Monthly\ \ Data = Seasonal + Trend + Remainder$$

- Two time series can appear correlated but after removing the seasonal component the correlation vanishes

Correlation plot original SST

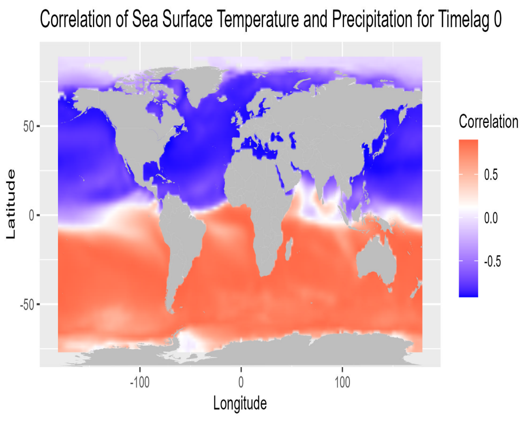


Figure 7: Correlation plot between SST and mean precipitation in the CAB for timelag 0

Correlation plot original SST

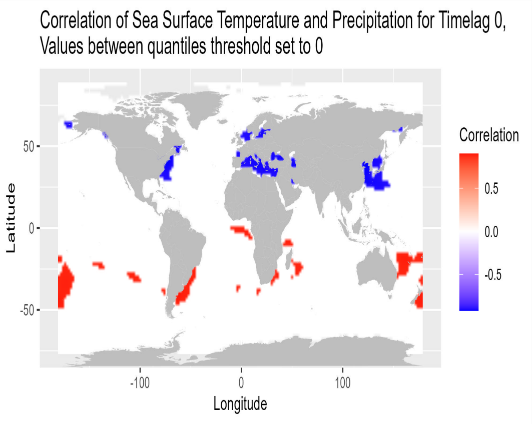


Figure 8: Correlation plot between SST and mean precipitation in the CAB for timelag 0

Correlation plot de-seasonalized SST

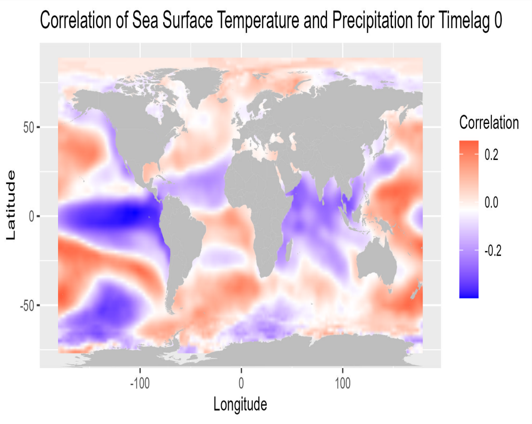


Figure 9: Correlation plot between de-seasonalized SST and de-seasonalized mean precipitation in the CAB for timelag 0

Correlation plot de-seasonalized SST

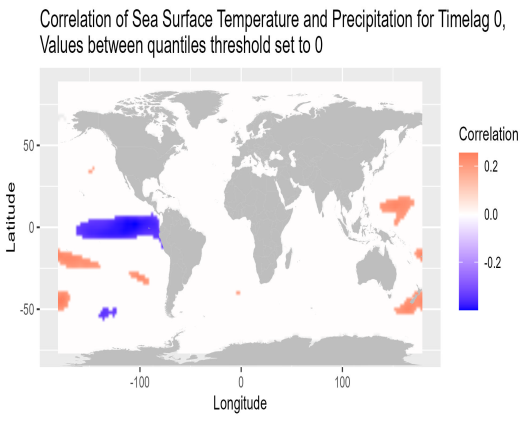


Figure 10: Correlation plot between de-seasonalized SST and de-seasonalized mean precipitation in the CAB for timelag 0

Clustering

Motivation/ Overview

- explorative analysis has shown spatial and temporal differences in the precipitation data
- we explored this further using k-means clustering
- steps: find optimal k via pca and gap statistic
- apply k-means to original precipitation data
- we compared k-means and k-medoid with and without PCA via the gap statistic
- here show only k-means with PCA as it gave best results
- applying the regression models to separate clusters might improve predictions
- Using 3 principal components and 5 cluster centers with k-means gave best results on gap statistic

k-means

- Our objective is to find k internally homogeneous and externally heterogeneous clusters
- Similarity is measured by the euclidean distance

$$d(x_i, x_{i'}) = \sum_{j=1}^{p}(x_{ij} - x_{i'j})^2 = ||x_i - x_{i'}||^2 (\#eq:eucl-dist) \quad (1)$$

- And we want to minimize the sum of distances inside all clusters, given by:

$$W(C) = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\sum_{C(i')=k}||x_i - x_{i'}||^2 = \sum_{k=1}^{K} N_k \sum_{C(i)=k}||x_i - \bar{x}_k||^2 (\#eq:w \quad (2)$$

where $\bar{x} = (\bar{x}_{1k}, ..., \bar{x}_{pk})$ stands for the mean vectors of the kth

gap statistic

- number of clusters has to be defined beforehand
- we decided on the optimal number of k using the gap statistic
- Let W_k be $W(C)$ for fix k
- We compare W_k from the precipitation data with average W_k^* from B Monte Carlo sampled data sets

$$Gap(k) = E\{log(W*_k)\} - log(W_k).(\#eq : gap) \tag{3}$$

- We choose k as smallest k such that

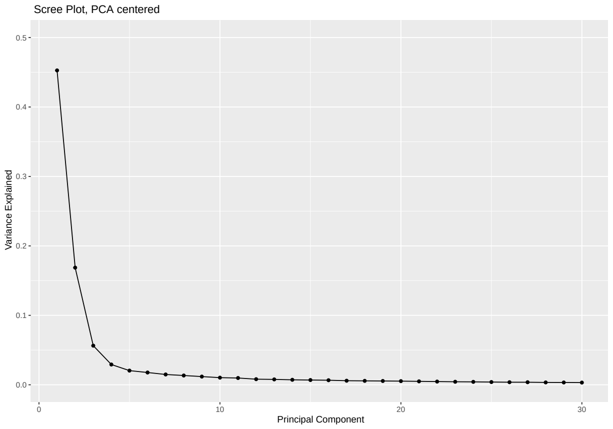
$$Gap(k) \geq Gap(k+1) - s_{k+1} \tag{4}$$

- s_{k+1} is $sd_k\sqrt{1+1/B}$, and sd the standard deviation of log(W*_k)

PCA

- Before running k-means we center the precipitation data and apply a PCA to reduce the large number of correlated variables to a few
- The new variables are linear combinations of the original variables
- Here: Each variable is a month of precipitation data in the CAB

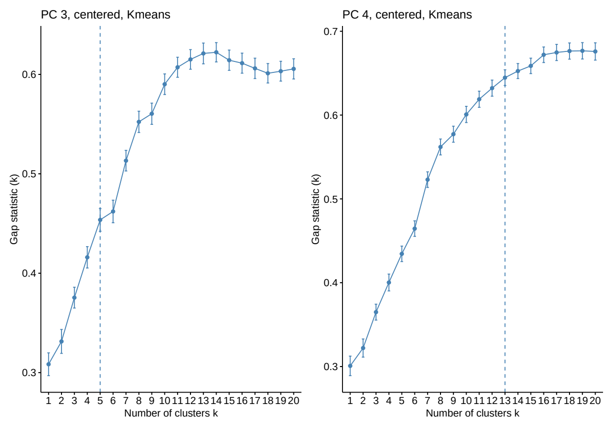
Scree Plot, PCA after centering



Screeplot

- The “elbow” be observe in the screeplot suggest 3 or 4 principal components
- The first 3 and 4 first PC explain 67.77 and 70.79 of the variance respectively.
- We compare the gap statistic results for 3 and 4 PC

Gap statistic results



- The k-means gap statistic on the first 3 PC proposes 5 clusters - For 4 PC, 13 clusters are chosen
- We chose 5 clusters since the result on 3 PC appears to be clearer and 5 clusters are more applicable than fitting the

Clustering results

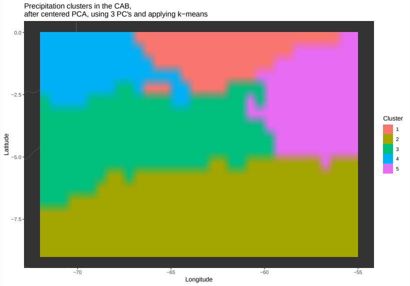


Figure 11: Spatial distribution of the found clusters in the CAB. We applied a centered PCA on the data and used 3 principal components before applying the k-means algorithm

- We find 5 clusters of different sizes
- The found clusters are almost completely spatially coherent although we did not include any spatial dependencies in the clustering

The lasso

Definition of the lasso

- We now consider the lasso regression problem

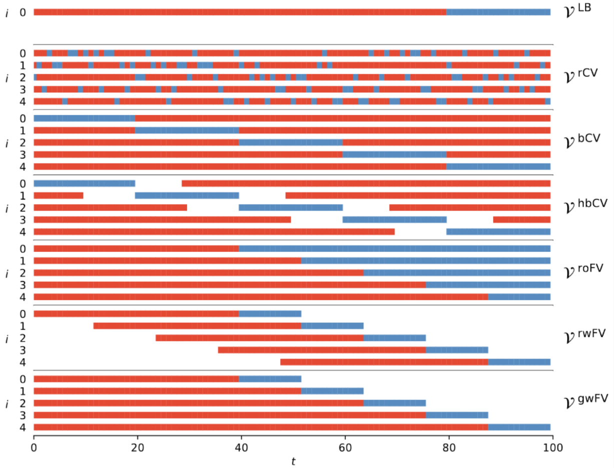
$$\min_{\beta_0,\beta} \frac{1}{N} \sum_{i=1}^{N} w_i l(y_i, \beta_0 + \beta^T x_i) + \lambda[(1-\alpha)||\beta||_2^2/2 + \alpha||\beta||_1](\#eq:glmnet) \tag{5}$$

- In our setting n « p, so lasso is natural choice
- The problem is solved using coordinate descent
- Due to the time dependencies in our data normal Cross Validation may be unjustified

Model evaluation

- Our goal is to train a model that can also predict well on new, unseen data
- We simulate the situation of unseen data by splitting our data into one part for model selection and another part for model evaluation
- Model evaluation is usually done via Cross Validation, but classic Cross Validation does not take into account the time dependency in our data

Forward selection



Forward selection

- We compute a λ-vector for the complete training set
- For each fold we fit a model with this λ-vector
- We compute the prediction error for the cv-test set of each fold
- Choose λ_{min}, λ that minimizes average MSE over all folds
- Fit model on complete selection data with λ_{min} and compute MSE on evaluation data

lasso settings

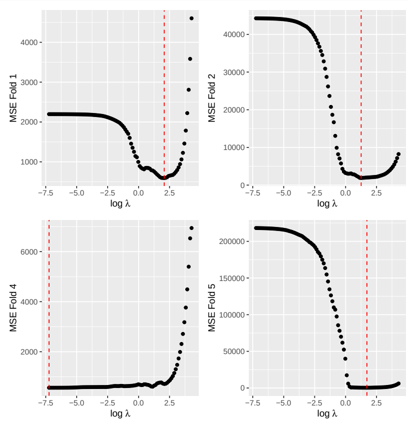
- lasso
- lasso with standardized features
- lasso with de-seasonalized SST
- lasso with differentiated SST
- lasso on clusters

lasso results TODO

- Show only best model results
- lasso with standardized features
- show MSE in plots
- show predictions in plots
- show predictions
- show coefficients
- display table

Lasso results

MSE in each fold



	min err	log lambda
Fold 1	594.04	2.07
Fold 2	1887.91	1.26
Fold 3	785.07	0.21
Fold 4	561.13	−7.23
Fold 5	347.18	1.72
Best	917.28	1.26

Predictions on test set, for each Fold

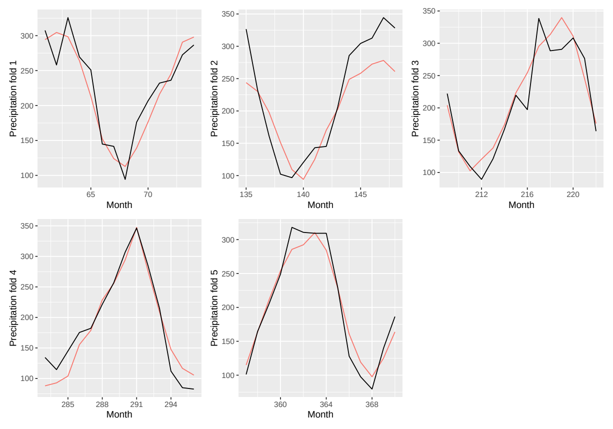


Figure 12: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set

Predictions on External Test Set

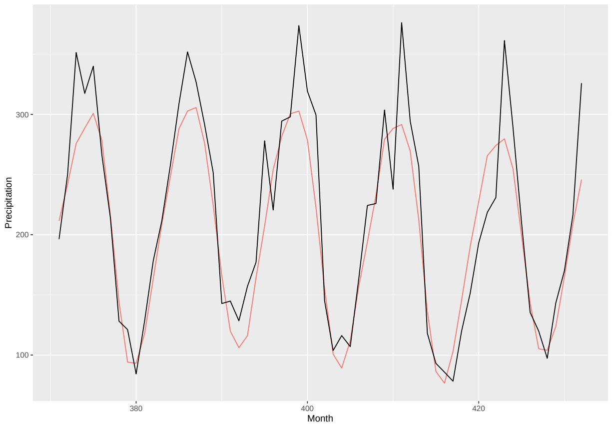


Figure 13: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

SST Regions chosen by the lasso

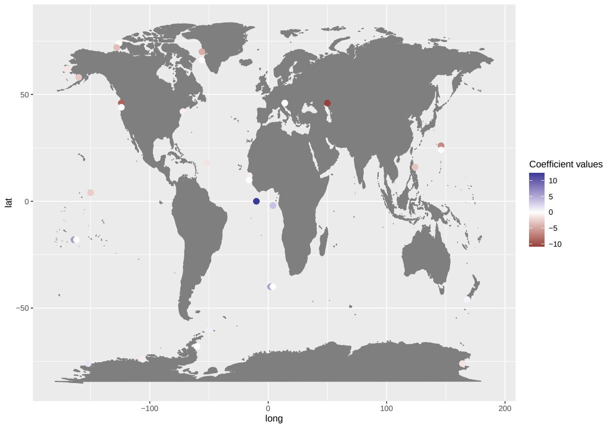


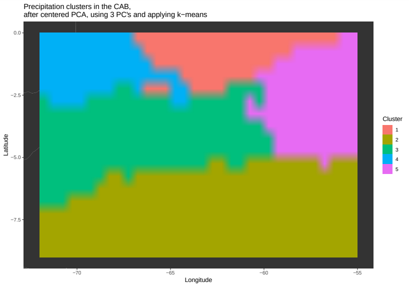
Figure 14: Coefficient plot of the full lasso model.

Lasso results all models

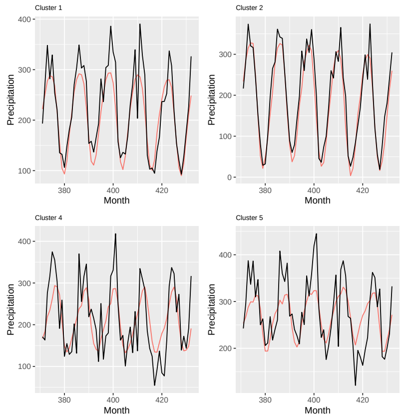
Lasso results all models

	MSE	Lambda
Standardized	1214.49	3.52
Original	1314.93	3.52
Differentiated	1361.82	2.21
De-seasonalized	1809.45	1.75

Lasso on clusters results



MSE clustered results



	MSE, Lasso
Cluster 1	1616.59
Cluster 2	1250.8
Cluster 3	3091.45
Cluster 4	3505.33
Cluster 5	2599.92
No clustering	1314.93

Summary lasso results

- We compared different settings for the lasso
- Lasso, lasso with standardization, de-seasonalizing, differentiation and on clusters
- Lasso with standardized SST worked best
- Can predict general seasonality, but still fails to predict peaks in precipitation
- Clustering the CAB improves only on one cluster, but on this cluster peaks can be predicted better than in the unclustered model
- Lasso chooses single “points” and not whole areas
- The points chosen as coefficients differ in the models, and can be very far away from the CAB

Discussion Validation approach (maybe discuss this at the end)

- For the CAB we can not predict large values in the hold-out set, on cluster 2 it works a little better
- Possible explanations:
- Our validation approach works better when train and test set are similar in terms of seasonality and trend
- When train and test differ, predictions might not work so well (test of stationarity in folds)
- Differentiating and de-seasonalizing could not solve this problem
- Predictions work better when the precipitation remains fairly stable over time, see Cluster 2
- Final model uses complete model selection data, possibly some of that information is not useful anymore if it's toocfar away from hold-out time frame
- Our validation approach is a trade-off between efficient use of

The fused lasso

Definition of the fused lasso

- Fused lasso, "fuses" predictors together
- It penalizes the difference of close predictores
- Therefore close predictors should be similar

$$\min_{\beta} 1/2 \sum_{i=1}^{n} (y_i - x_i^T \beta_i)^2 + \lambda \sum_{i,j \in E} |\beta_i - \beta_j| + \gamma \cdot \lambda \sum_{i=1}^{p} |\beta_i|, (\#eq: fused-lasso) \tag{6}$$

- with x_i being the ith row of the predictor matrix and E is the edge set of an underlying graph.
- The third term $\gamma \cdot \lambda \sum_{i=1}^{p} |\beta_i|$, controls the sparsity of the coefficients.
- $\gamma = 0$ leads to complete fusion of the coefficients (no sparsity) and $\gamma > 0$ introduces sparsity to the solution, with higher values placing more priority on sparsity.

Fused lasso optimization

- Lets consider the problem in the notation of the generalized lasso problem

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|D\beta\|_1, (\#eq : gen - lasso) \qquad (7)$$

- where $y \in \mathbb{R}^n$ is the vector of the outcome, $X \in \mathbb{R}^{n \times p}$ a predictor matrix, $D \in \mathbb{R}^{m \times p}$ denotes a penalty matrix, and $\lambda \geq 0$ is a regularization parameter.
- The dual path algorithm solves not the primal but the dual solution of the problem and computes the solution for a whole path instead of single values of λ.

- Let's consider the case when $X = I$ and $rank(X) = p$ (this is called the "signal approximator" case), the dual problem of @ref(eq:gen-lasso) is then:

$$\hat{u} \in \underset{u \in \mathbb{R}^\omega}{\arg\min} \frac{1}{2} \left\| y - D^T u \right\|_{\frac{2}{2}} \text{ subject to } \|u\|_\infty \leq \lambda.(\#eq : dual) \tag{8}$$

- The primal and dual solutions, $\hat{\beta}$ and $\hat{u}$ are related by:

$$\hat{\beta} = y - D^T \hat{u}.(\#eq : dual - relate) \tag{9}$$

- For general X and D with exploitable structure (as in our case), specialized implementations exist

Graph structure

- We can use a graph as input in the fusedlasso function
- We created a grid and deleted all nodes that were land regions
- This induced subgraphs

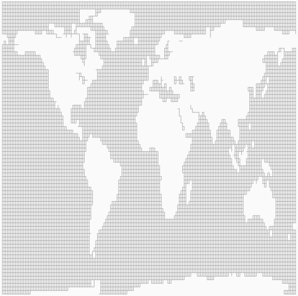


Figure 15: Graph of the SST and land areas used in fused lasso

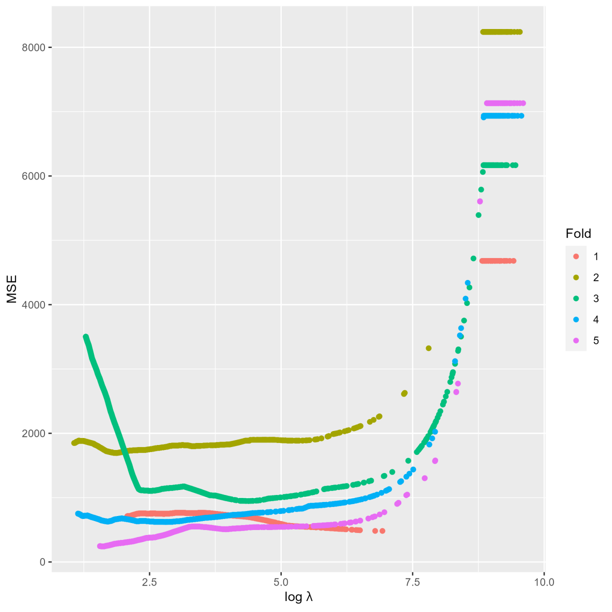
Graph structure and implications

- Results showed that removing the sub-graphs improved performance, although some of the regions were included in the final lasso models
- If we don't remove the clusters and also add sparsity (i.e $\gamma > 0$) the clusters dominate the results even more
- Possible explanations: Sub-graphs are less penalized, because they have fewer edges.
- Removing the clusters improved results more than f.e standardization

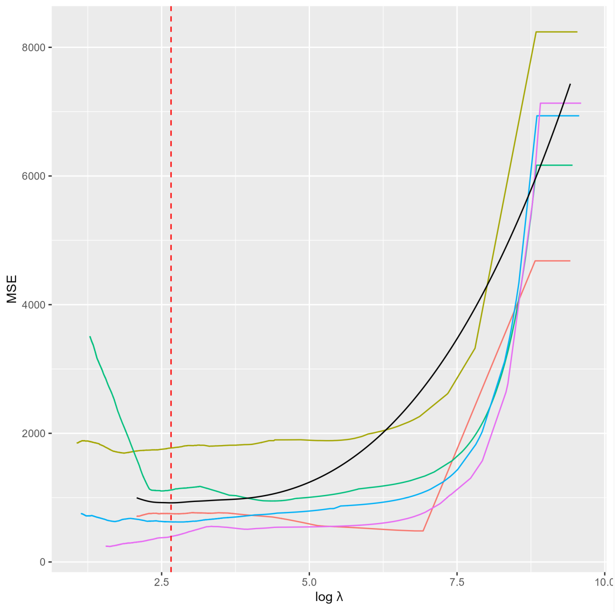
Fused lasso settings

- The considered fused lasso settings are: Fused lasso with clusters, fused lasso without clusters, fused lasso without clusters and sparsity (gamma: 0.01, 0.05, 0.1)
- Fused lasso without clusters and no sparsity showed best results

Fused lasso results, clusters removed



Fused lasso results, clusters removed



Prediction plots

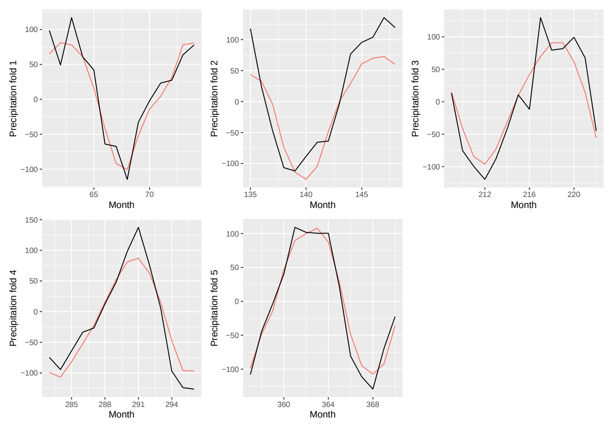


Figure 16: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Prediction plot

The predictions inside the folds are very similar to lasso without standardization, the same holds for the predictions from the full model, but the MSE improves here.

Full predictions

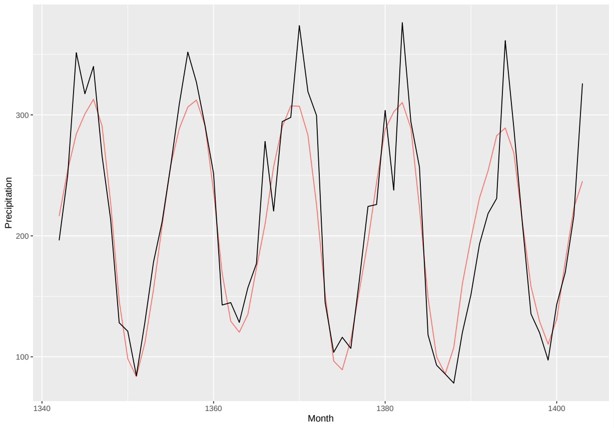
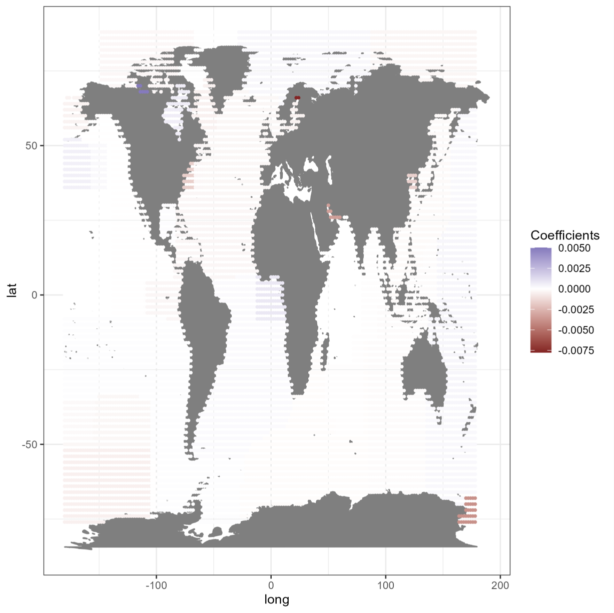
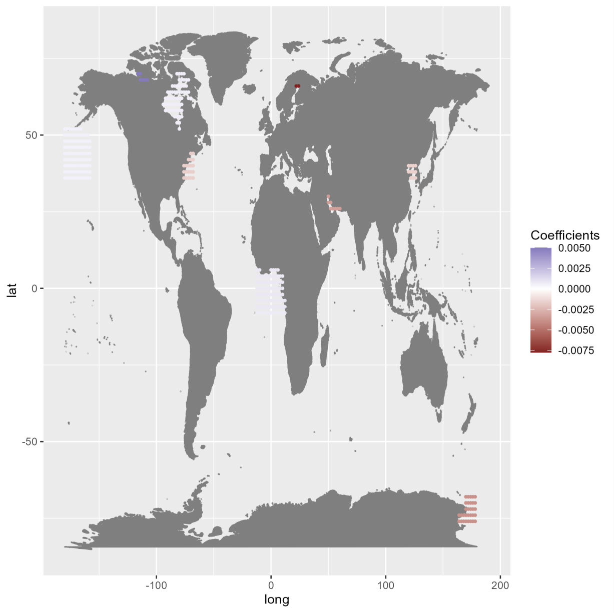


Figure 17: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

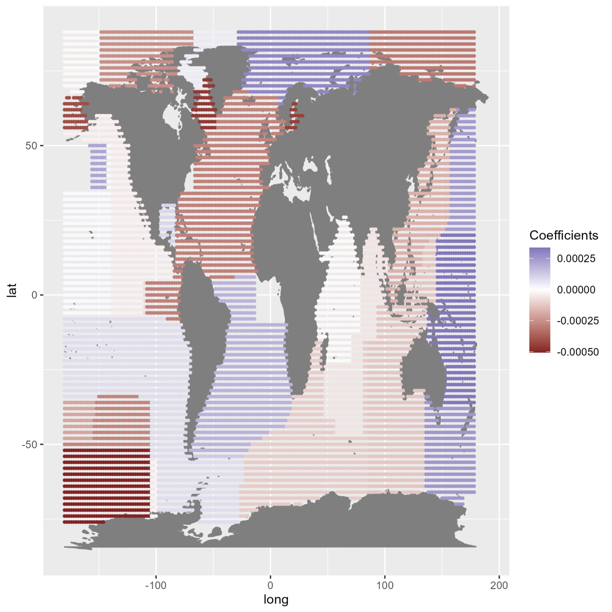
Coefficient plot



Coefficient plot, highest absolute values only



Coefficient plot, lowest absolute values only



Fused lasso results MSE

	MSE	Lambda
No sub-graphs	1070.04	14.28
With sub-graphs	1131.71	18.55
No sub-graphs, gamma 0.05	1836.63	2544.23
No sub-graphs, gamma 0.1	1840.59	1586.29

Fused lasso results summary

- We compared different settings for the fused lasso, removing the sub-graphs and introducing no sparsity gave the best results
- Removing the sub-graphs removed some of the optimization problems, but nodes with less edges are still less penalized
- Implementing a validation strategy was more complex than for the lasso
- We smoothed the error-lines in each fold over a common region to compute λ_{min}
- The coefficient plots reveal predictive areas with high negative values in the Baltic Sea and high positive values north east of Canada.
- Since no sparsity is used, all areas obtain non-zero coefficient values

Fused lasso discussion

- Computing the solution path is computationally expensive
- The graph structure is highly influential and cost will scale with number of edges
- While the best fused lasso approach performed best overall, it still is not able to predict high precipitation values
- Possible improvements on optimization path: creating weighted graph (increases number of edges and cost), narrowing down the SST "window" (f.e as in Ciemer et al. (2020))
- Possible improvements on feature engineering: as for the lasso, differentiating, de-seasonalizing

Discussion & Conclusion

Discussion

- Our results suggest that precipitation can to some extend be predicted from SST directly.
- The overall predictability of precipitation in the CAB differed between model selection and model evaluation phase.
- For one part this might be due to the difference in the regions in the CAB, since clustering improved the results for one specific cluster.
- Another explanation could be that our model selection approach was not optimal in its use of the data.
- We might have been to restrictive in exploiting the data or used data that became less relevant over time.

Discussion

- Possible other approaches:
- Allow for larger folds (introduces overlapping folds), or for crossing of train and test in time (past is predicted with future values)
- Fit the full model with less data and discarding data that is far away from the hold-out validation set.
- The results of the fused lasso will depend a lot on the graph structure, sub-graphs do not represent the real situation well
- Creating a weighted graph or narrowing the SST “window” may improve performance.
- Also, applying the fused lasso only on the best performing cluster from the lasso may yield better results.

Conclusion

- In a descriptive analysis we found temporal and spatial patterns in the correlation of rain in the CAB and SST
- The cluster analysis revealed 5 almost completely spatially coherent clusters in the CAB
- Standardizing the features yielded the best results for the lasso
- The lasso can predict the precipitation on the model selection test sets a lot better than in the hold-out test set
- On the hold-out data the lasso fails to predict the peaks in precipitation

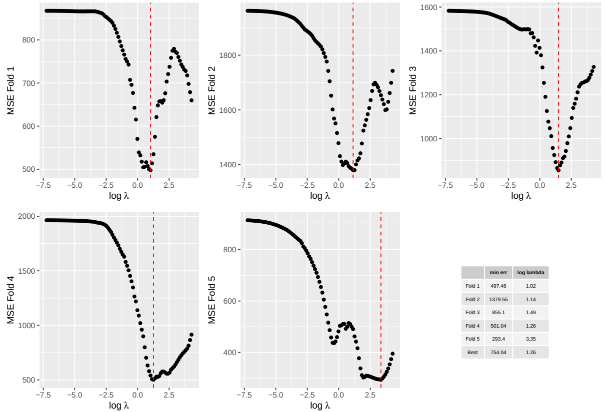
Conclusion

- We applied the fused lasso to our problem and implemented a model evaluation approach
- The fused lasso improves predictive power compared to the lasso when the sub-graphs are removed
- The fused lasso is still is not able to predict high values in precipitation well
- We could further improve the clustering method by taking into account spatial dependencies.
- The fused lasso could be improved by using other model selection approaches or increasing the complexity of the graph structure.

Appendix

Lasso on original SST

MSE in each fold (Lasso on original SST)



	min err	log lambda
Fold 1	497.46	1.02
Fold 2	1379.55	1.14
Fold 3	855.1	1.49
Fold 4	501.04	1.26
Fold 5	293.4	3.35
Best	754.04	1.26

Predictions on test set, for each Fold (Lasso original SST)

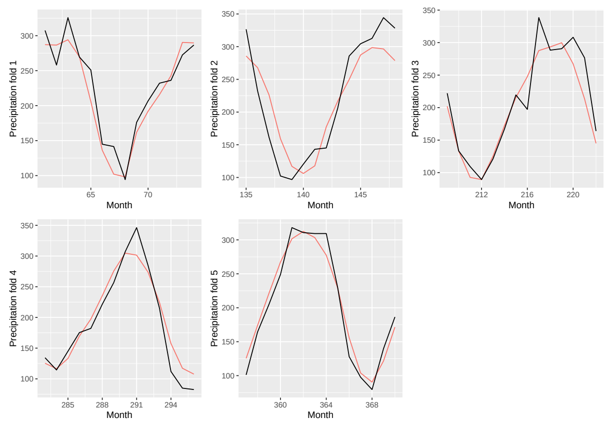


Figure 18: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set (Lasso on original SST)

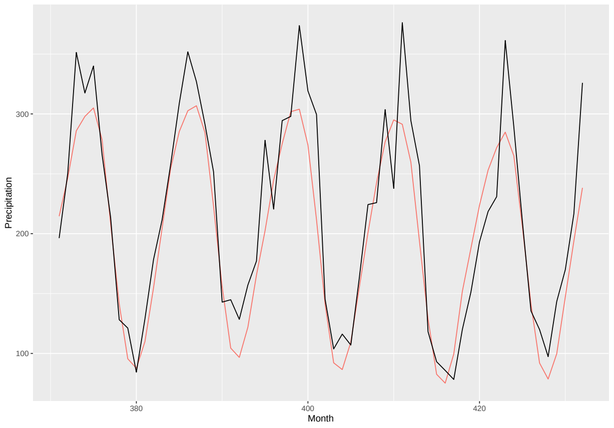


Figure 19: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

SST Regions chosen by the lasso (Lasso on original SST)

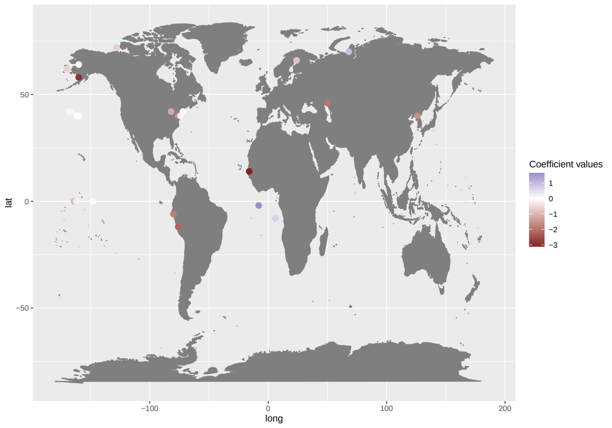
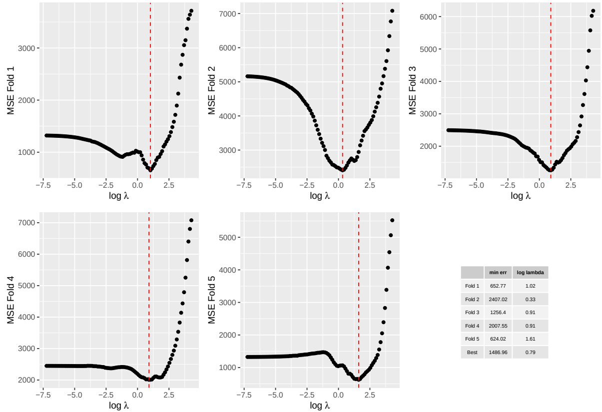


Figure 20: Coefficient plot of the full lasso model.

Lasso on differentiated SST

MSE in each fold (Lasso differentiated)



	min err	log lambda
Fold 1	652.77	1.02
Fold 2	2407.02	0.33
Fold 3	1256.4	0.91
Fold 4	2007.55	0.91
Fold 5	624.02	1.61
Best	1486.96	0.79

Predictions on test set, for each Fold (Lasso differentiated)

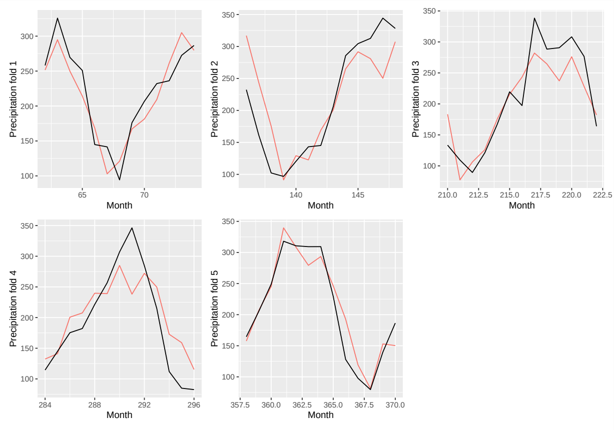


Figure 21: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set (Lasso differentiated)

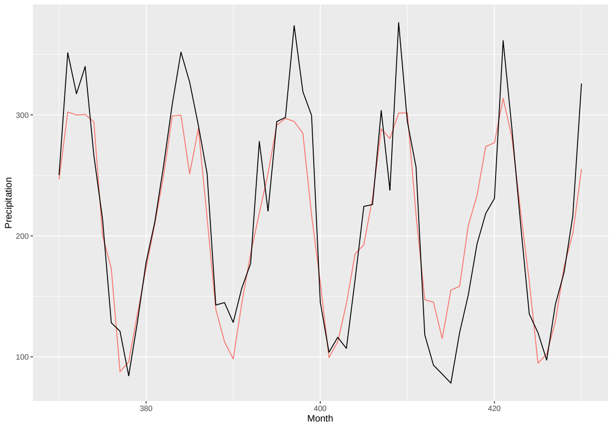


Figure 22: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

SST Regions chosen by the lasso (Lasso differentiated)

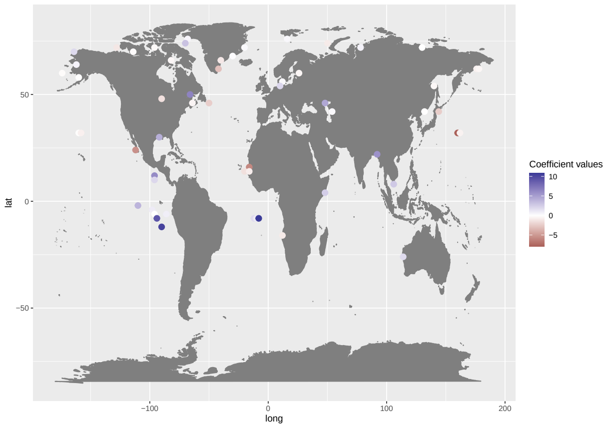


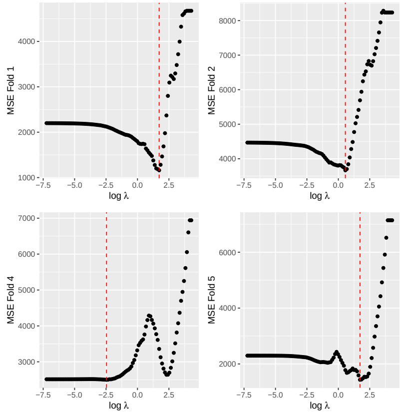
Figure 23: Coefficient plot of the full lasso model.

Fused evaluation (maybe explain this when showing resutls)

- Generally same setting as for lasso, 5 folds with train and test, choose λ_{min}, refit with λ_{min}, get MSE on hold-out test set.
- But for the fused lasso we can not define the λ vector beforehand.
- λ-path is found by dual path algorithm and the range of the paths can vary a lot!
- So to find λ_{min} we search over the common range of all folds and interpolate to lines
- λ_{min} is then the λ that minimize MSE over all λ of that common range

Lasso on de-seasonalized SST

MSE in each fold (Lasso de-seasonalized)



	min err	log lambda
Fold 1	1160.11	1.72
Fold 2	3665.68	0.56
Fold 3	967.65	0.56
Fold 4	2499.78	−2.46
Fold 5	1432.55	1.72
Best	2381.37	0.56

Predictions on test set, for each Fold (Lasso de-seasonalized)

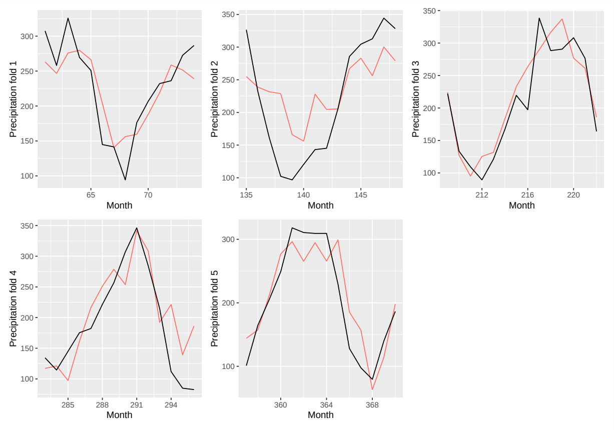


Figure 24: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set (Lasso differentiated)

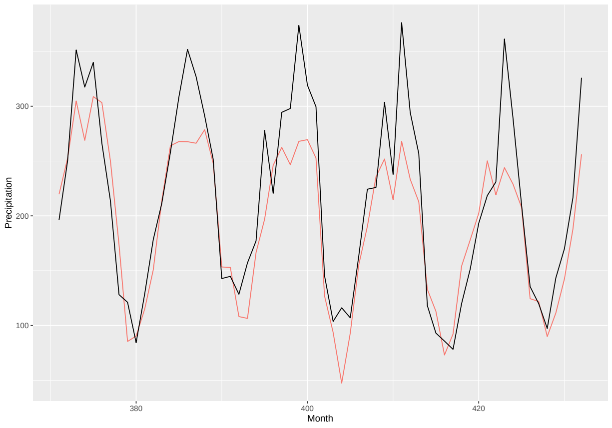


Figure 25: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

SST Regions chosen by the lasso (Lasso de-seasonalized)

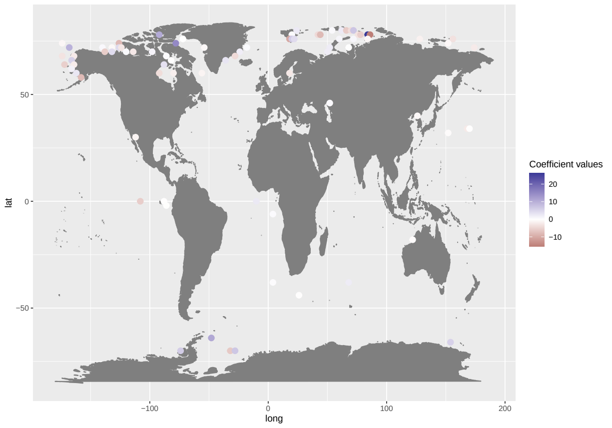
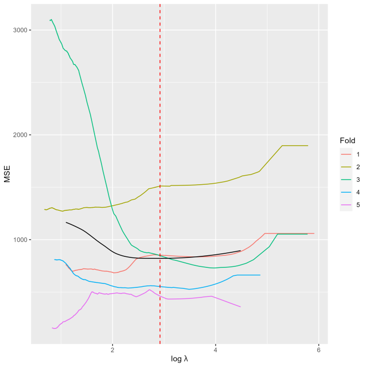


Figure 26: Coefficient plot of the full lasso model.

Fused lasso with sub-graphs

Error lines (Fused lasso with sub-graphs)



Prediction plots for each fold (Fused lasso with sub-graphs)

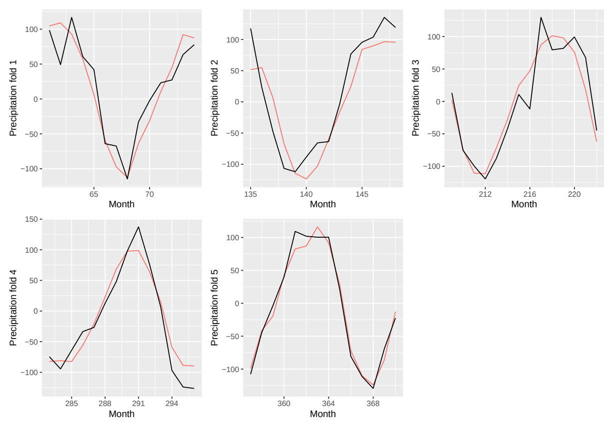


Figure 27: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

The predictions inside the folds are very similar to lasso without standardization (see @ref(fig:pred-plot-fold-lasso-og)), the same

Predictions on hold-out set (Fused lasso with sub-graphs)

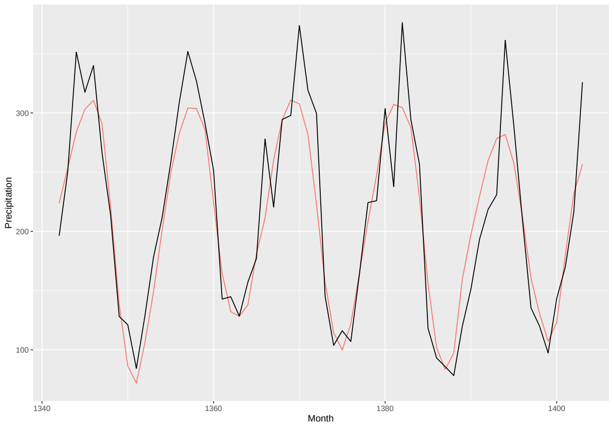
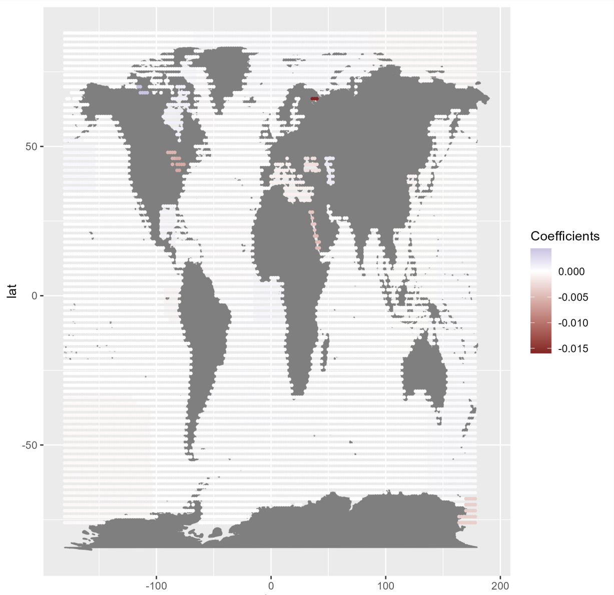


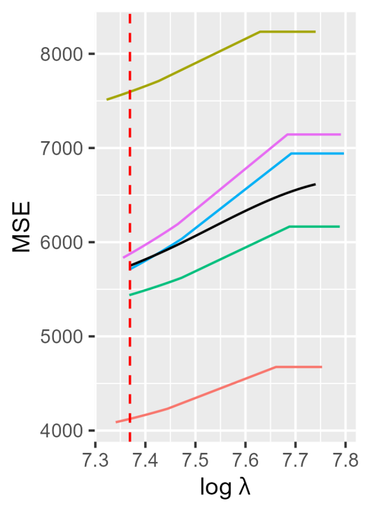
Figure 28: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

Coefficient plot of the final model (Fused lasso with sub-graphs)



Fused lasso without sub-graphs, gamma 0.1

Error lines (Fused lasso without sub-graphs, gamma 0.1)



Prediction plots for each fold (Fused lasso without sub-graphs, gamma 0.1)

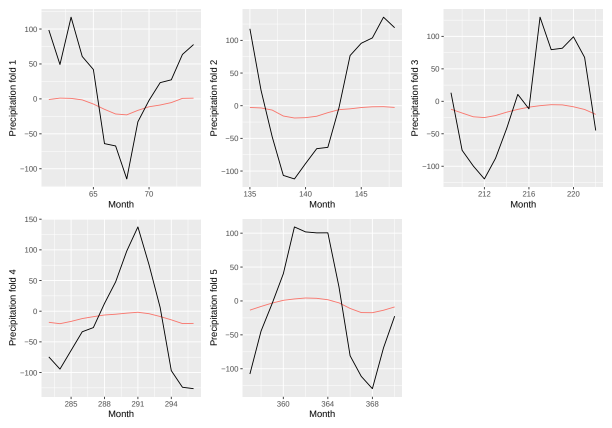


Figure 29: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

The predictions inside the folds are very similar to lasso without

Predictions on hold-out set (Fused lasso without sub-graphs, gamma 0.1)

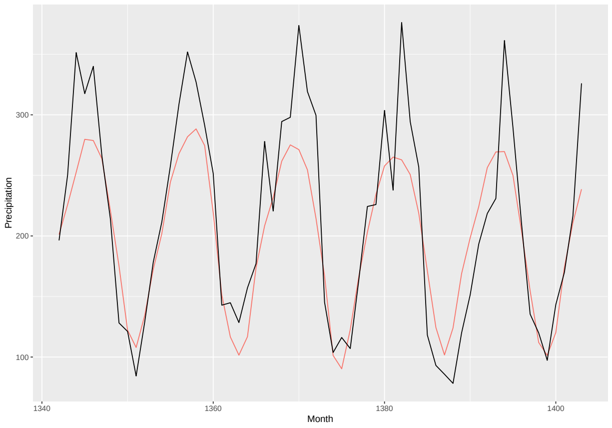
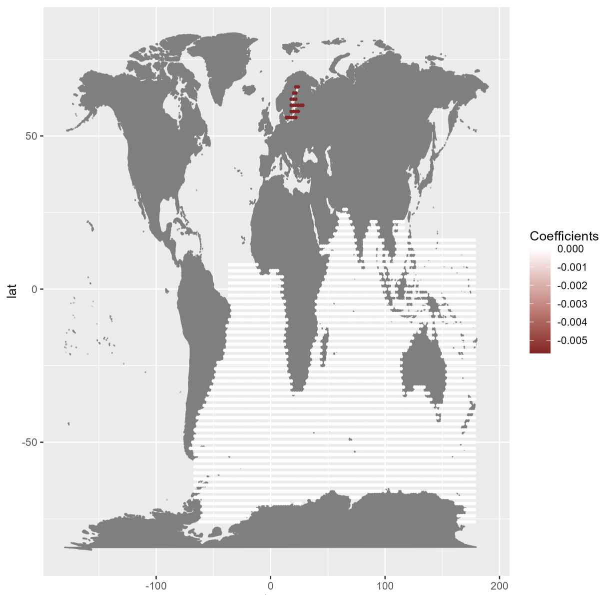


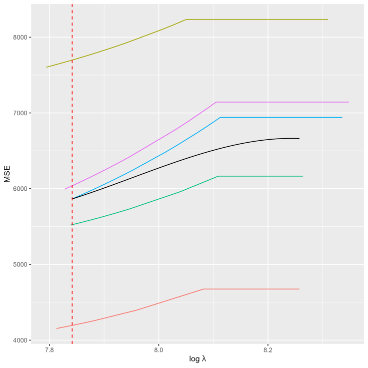
Figure 30: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

Coefficient plot of the final model (Fused lasso without subgraphs, gamma 0.1)



Fused lasso without sub-graphs, gamma 0.05

Error lines (Fused lasso without sub-graphs, gamma 0.05)



Prediction plots for each fold (Fused lasso without sub-graphs, gamma 0.05)

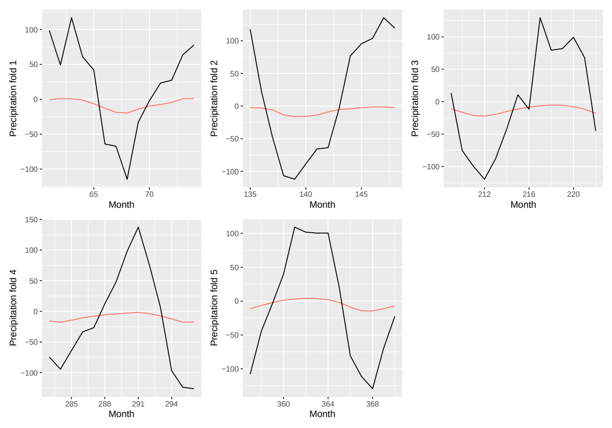


Figure 31: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

The predictions inside the folds are very similar to lasso without

Predictions on hold-out set (Fused lasso without sub-graphs, gamma 0.1)

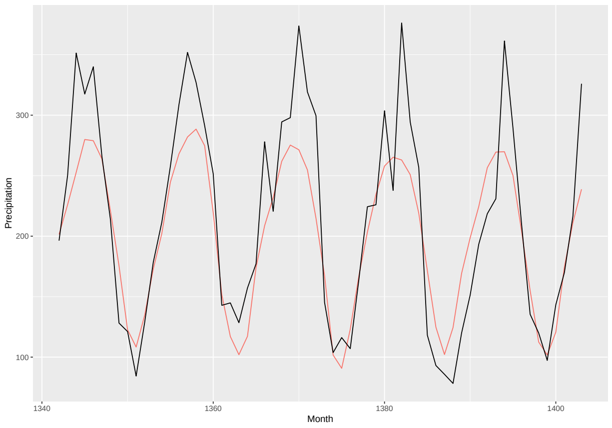
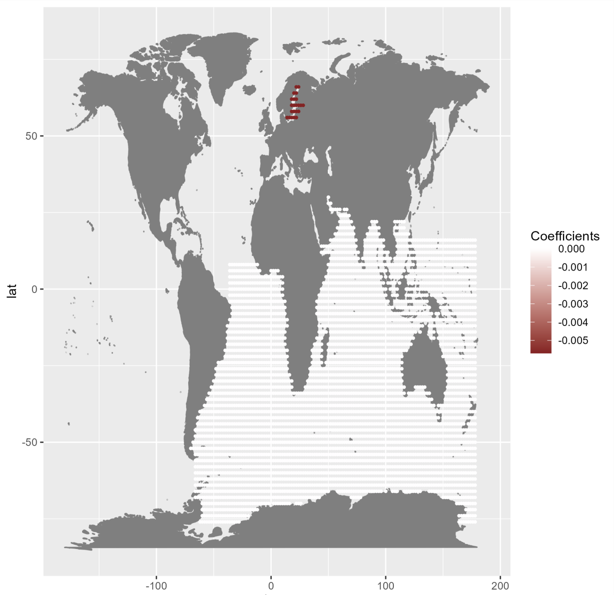


Figure 32: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

Coefficient plot of the final model (Fused lasso without subgraphs, gamma 0.05)



Fused evaluation (maybe explain this when showing resutls)

Fused evaluation (maybe explain this when showing resutls)

- Generally same setting as for lasso, 5 folds with train and test, choose λ_{min}, refit with λ_{min}, get MSE on hold-out test set.
- But for the fused lasso we can not define the λ vector beforehand.
- λ-path is found by dual path algorithm and the range of the paths can vary a lot!
- So to find λ_{min} we search over the common range of all folds and interpolate to lines
- λ_{min} is then the λ that minimize MSE over all λ of that common range

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.