

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

Author:

Dario Lepke

Supervisors:

Dr. Fabian Scheipl (LMU Munich)

Dr. Niklas Boers (PIC Potsdam)

September 08, 2022

Outline

1. Introduction
2. Explorative analysis
3. Correlation analysis
4. Clustering precipitation
5. The lasso
6. The fused lasso
7. Discussion & Conclusion

1. 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 mio km^2

Related work

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

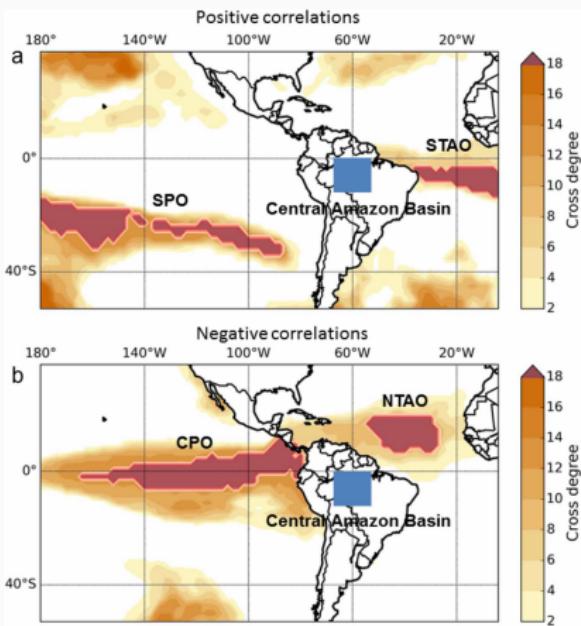


Figure 1: Cross degree between sea surface temperature and continental rainfall anomalies. (Ciemer et al. (2020))

Early warning signal

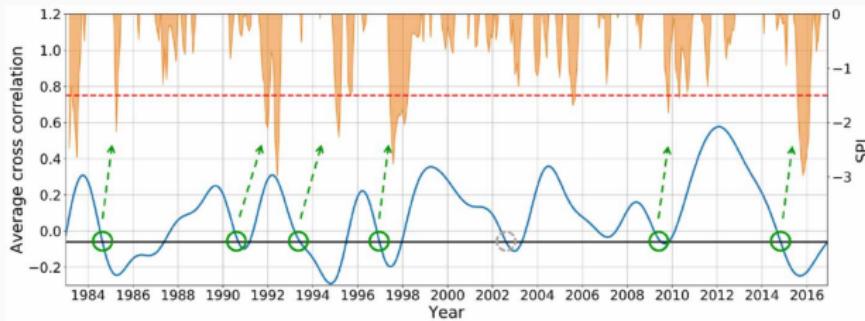


Figure 2: Early-warning signal for droughts in the central Amazon basin
Ciemer et al. (2020).

Our Approach

- Inspect spatial and temporal characteristics in the raw data
- Fit different models to predict precipitation directly from SST
- Apply the lasso and fused lasso with different parameter settings (RJ Tibshirani (2011), Robert Tibshirani et al. (2005))
- Evaluate the models with forward validation for time series

2. Explorative analysis

The Data

- Rain data from CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data, funk2015climate)
- CHIRPS derived from in-situ and satellite data
- SST data from ERSST (Extended Reconstructed Sea Surface Temperature, huang2017noaa)
- 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

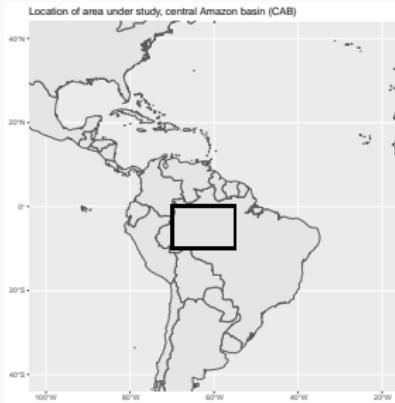


Figure 3: Location of the area under study. The central amazon basin (CAB) spanning across 0,-10 latitude and -70,-55 longitude

Precipitation, Mean and SD

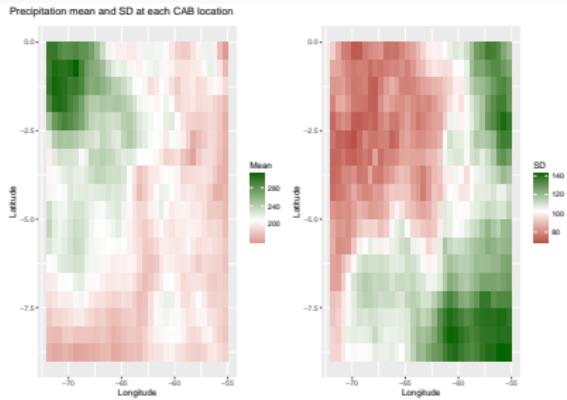


Figure 4: Mean and standard deviation at each location. The standard deviation was computed over the whole time period. The white lines on the legend at each side of the plots indicates the mean of the respective quantity.

Precipitation Glyph Plots

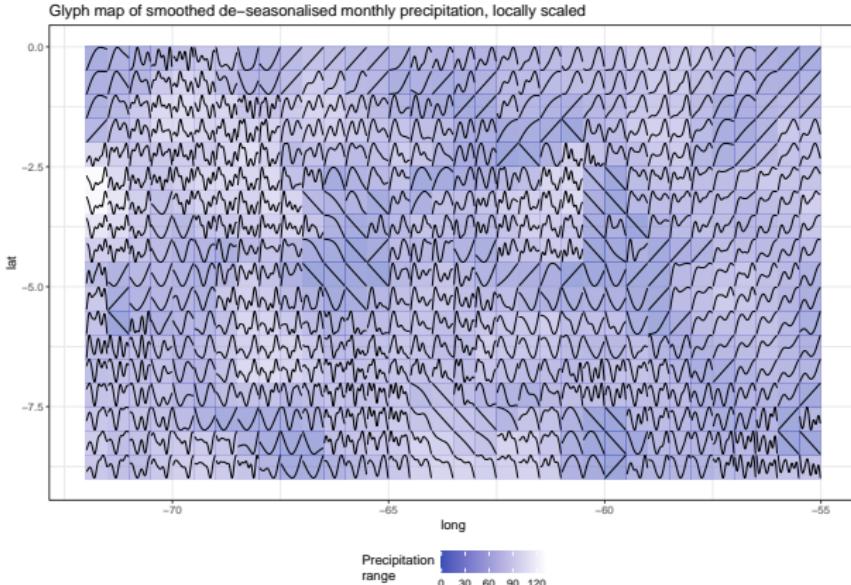


Figure 5: Glyph map of de-seasonalised and smoothed precipitation. The time series are scaled locally, ranges are not the same in all cells. The different ranges are given in color shades, where lighter shading indicates a larger range and darker shades smaller ranges.

Explorative analysis SST

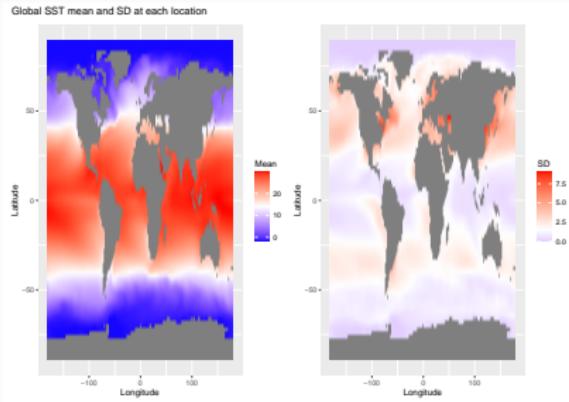


Figure 6: Mean and SD SST on the global map. The white lines on each legend at the side of the plots indicates the mean of the respective quantity.

3. Correlation analysis

Correlation analysis

- Here we get an overview over the general correlation structure of the data
- We show the correlations 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

$$\text{Monthly Data} = \text{Seasonal} + \text{Trend} + \text{Remainder}$$

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

Correlation plot original SST

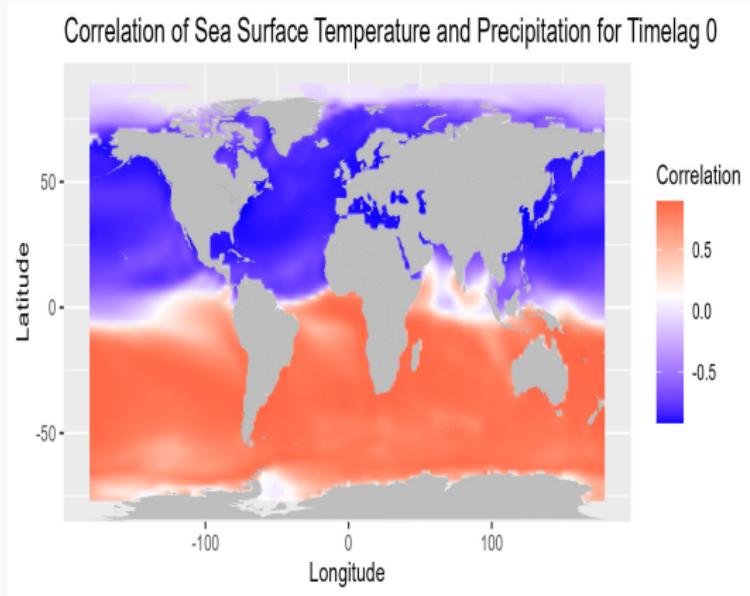


Figure 7: Correlation plot between SST and mean precipitation in the CAB for timelag 0.

Correlation plot original SST

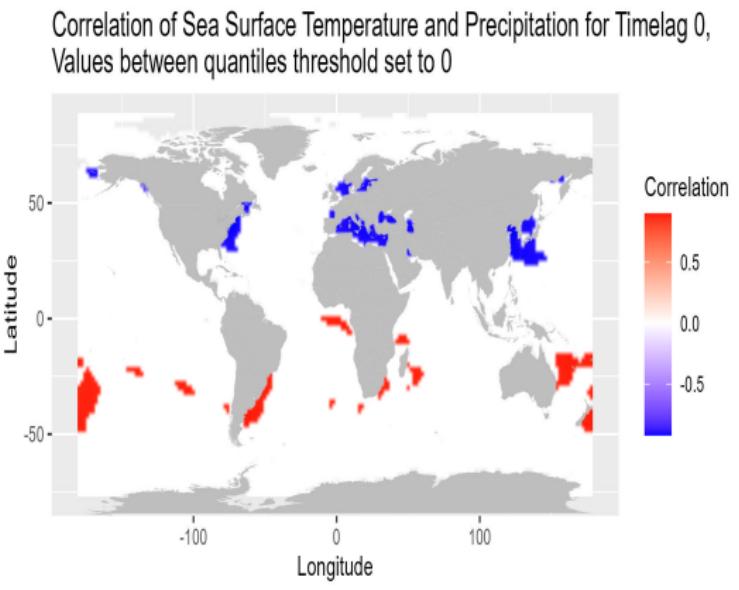


Figure 8: Correlation plot between SST and mean precipitation in the CAB for timelag 0. Values between 0.975 and 0.25 quantiles are set to 0.

Correlation plot de-seasonalized SST

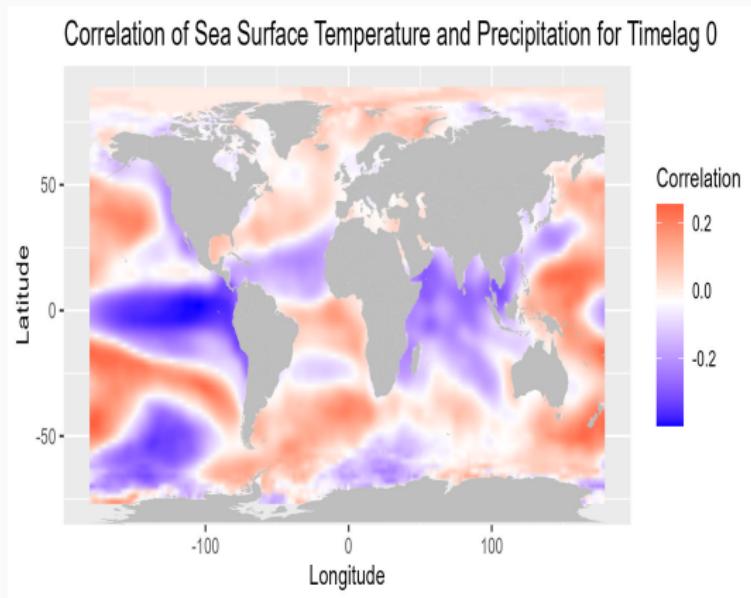


Figure 9: Correlation plot between de-seasonalized SST and de-seasonalized mean precipitation in the CAB for timelag 0.

Correlation plot de-seasonalized SST

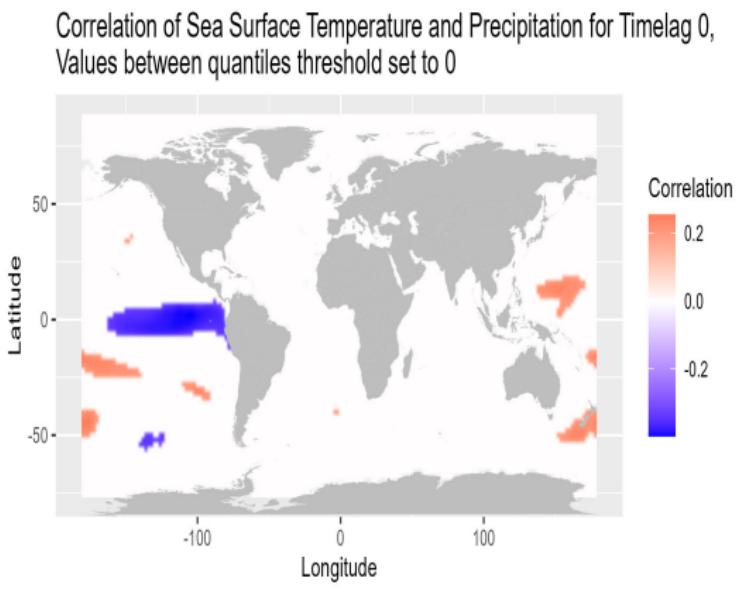


Figure 10: Correlation plot between de-seasonalized SST and de-seasonalized mean precipitation in the CAB for timelag 0

4. Clustering precipitation

Clustering Motivation

- Our explorative analysis has shown spatial and temporal differences in the precipitation data
- We explore this further using k -means clustering:
 - find optimal k via PCA and gap statistic
 - apply k-means to original precipitation data
- We compare k-means with and without PCA, via the gap statistic
- Ultimate goal is to improve predictions by applying a model to each cluster separately

k-means

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

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

k-means

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

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

- where $\bar{x} = (\bar{x}_{1k}, \dots, \bar{x}_{pk})$ stands for the mean vectors of the k -th cluster and $N_k = \sum_{i=1}^N I(C(i) = k)$.

Gap Statistic

- The number of clusters (i.e k) has to be defined beforehand
- 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

$$Gap(k) = E\{\log(W_k^*)\} - \log(W_k). \quad (3)$$

- We choose k as smallest k such that

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

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

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

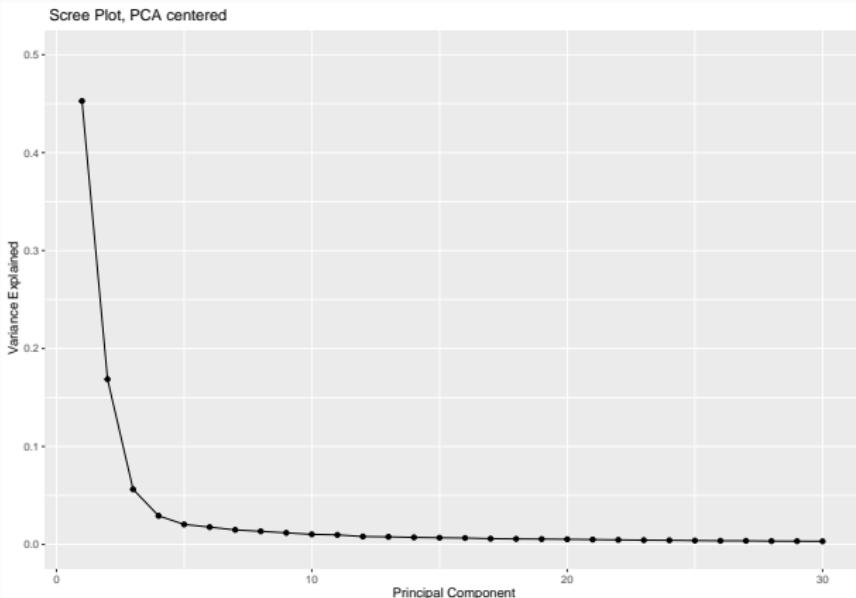


Figure 11: Scree plot of the Principal component analysis applied to the centered precipitation data.

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

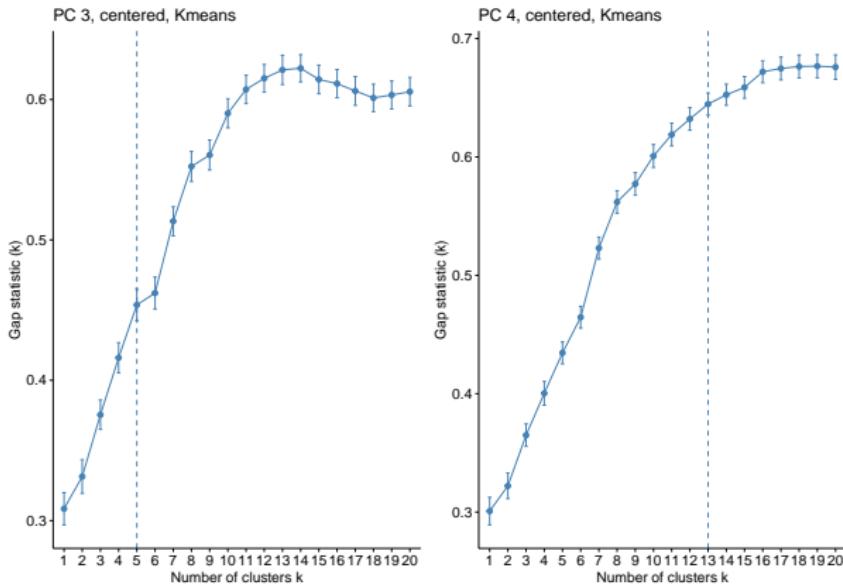


Figure 12: Results of the gap statistic when applying k-means on 3 (left) and 4 (right) principal components of the precipitation data.

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 model evaluation on 13 clusters.

Clustering results

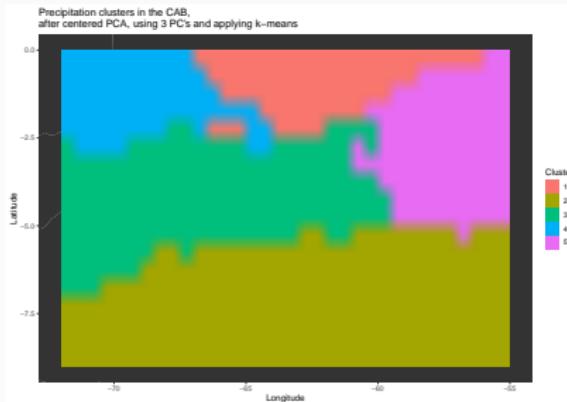


Figure 13: Spatial distribution of the found clusters in the CAB. We applied a centered PCA on the data and used 3 principal components before applying the k-means algorithm

- 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

Clustering results

- Small exception is the “island” of cluster 1 (orange) inside cluster 4 (blue) and on the edge on cluster 3 (green)
- Usefulness of clustering can only be determined after model fitting on each cluster

5. The lasso

Definition of the lasso

- In our setting $n \ll p$, so the lasso is a natural choice
- We consider the lasso regression problem:

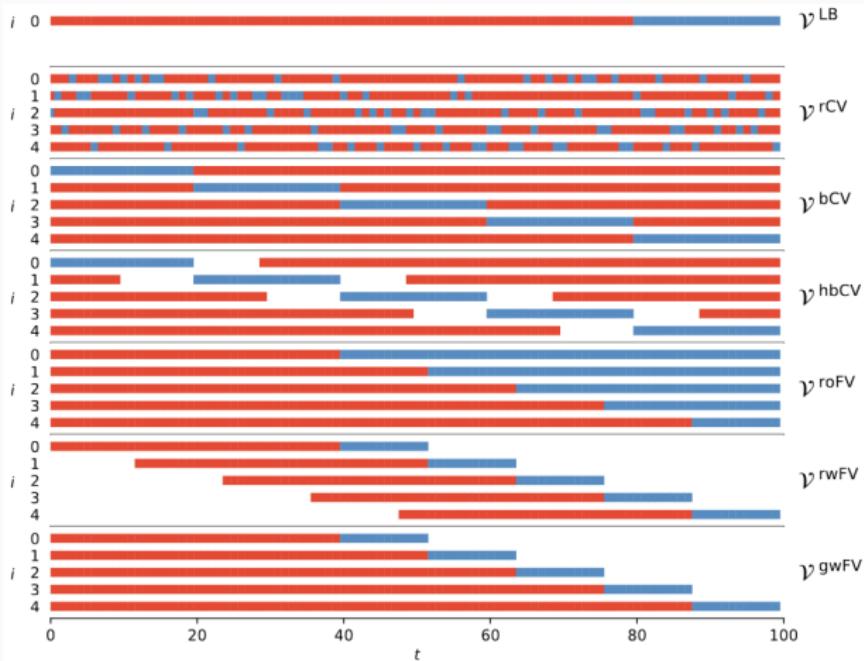
$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N l(y_i, \beta_0 + \beta^T x_i) + \lambda \|\beta\|_1 \quad (5)$$

- The problem is solved using coordinate descent (Van der Kooij (2007))
- Due to the time dependencies in our data normal Cross Validation may be unjustified

Model validation

- To find the best λ , we can usually use cross validation, to achieve a Bias-Variance trade-off.
- Classic cross validation uses the data efficiently, but we want to avoid predicting past values learned from future data.
- We don't want to predict past precipitation with models that learned on future data.

Forward validation



Forward validation

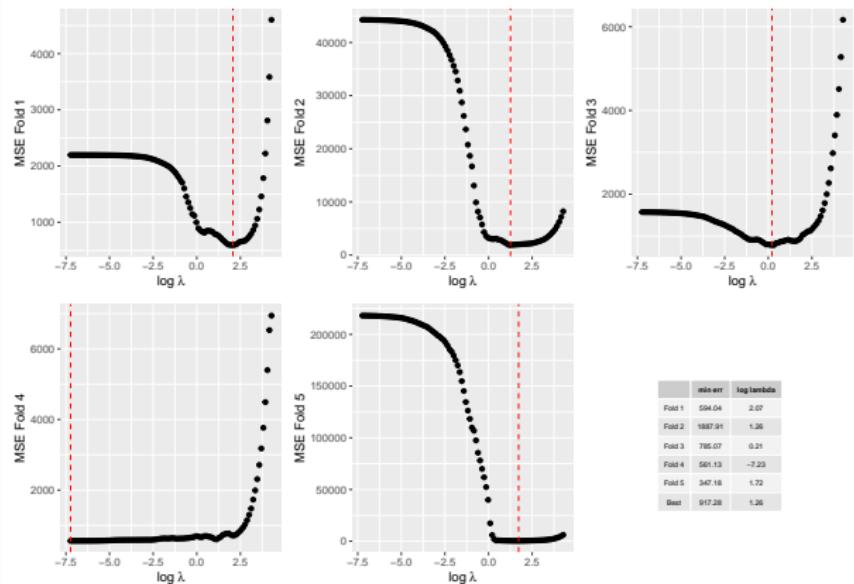
- We also want to ensure that each data point is only included once as training or test
- The choice is then to use non-overlapping rolling window forward validation with 5 folds
- We also use some part of the data as a hold-out test set that is not used in forward validation.
- The model selection set are the first 360 months, and the hold-out test set consists of the months 371 to 432 (62 months).
- Each fold has 60 months of training and 14 months of test.

Forward selection

- We compute a λ -vector for the complete model selection set
- For each fold we fit models with this λ -vector
- We compute the prediction errors on each FV-test set, for 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 results

MSE in each fold



Predictions on test set, for each Fold

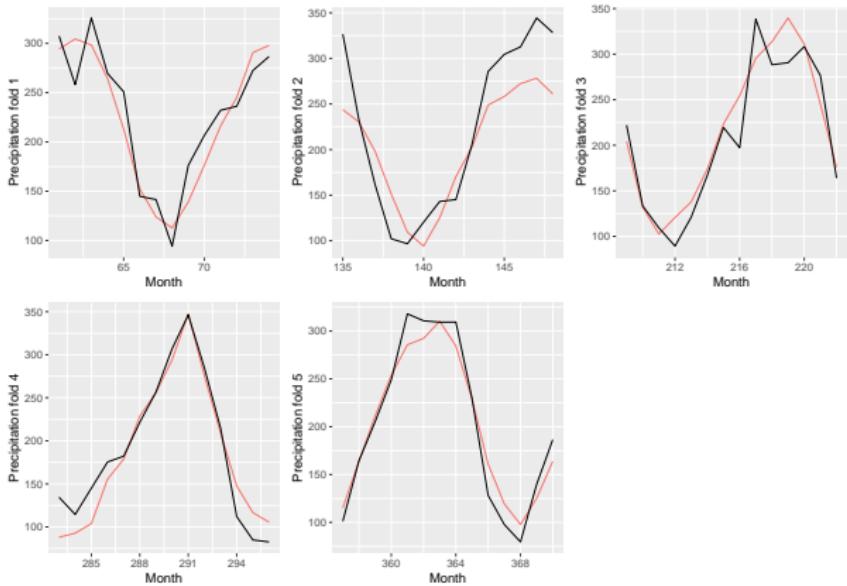


Figure 14: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set

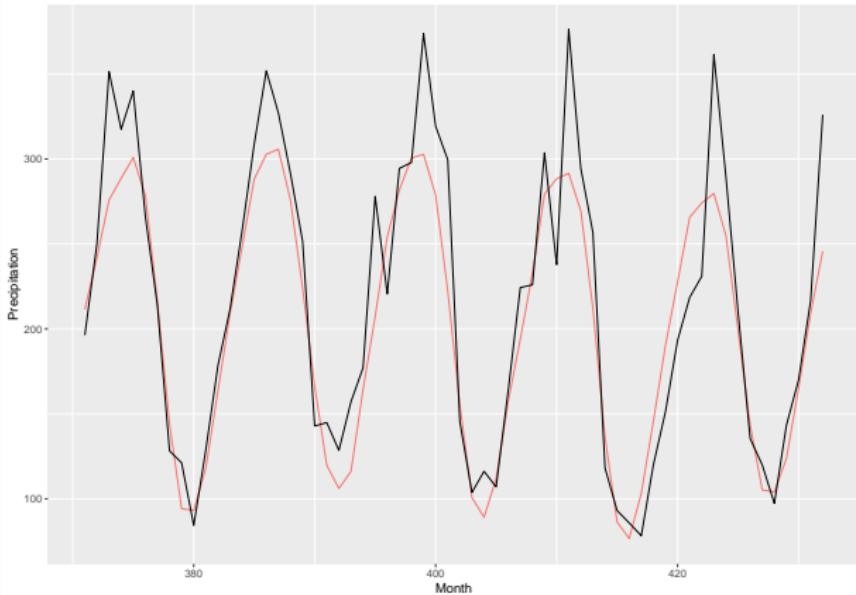


Figure 15: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

SST Regions chosen by the lasso

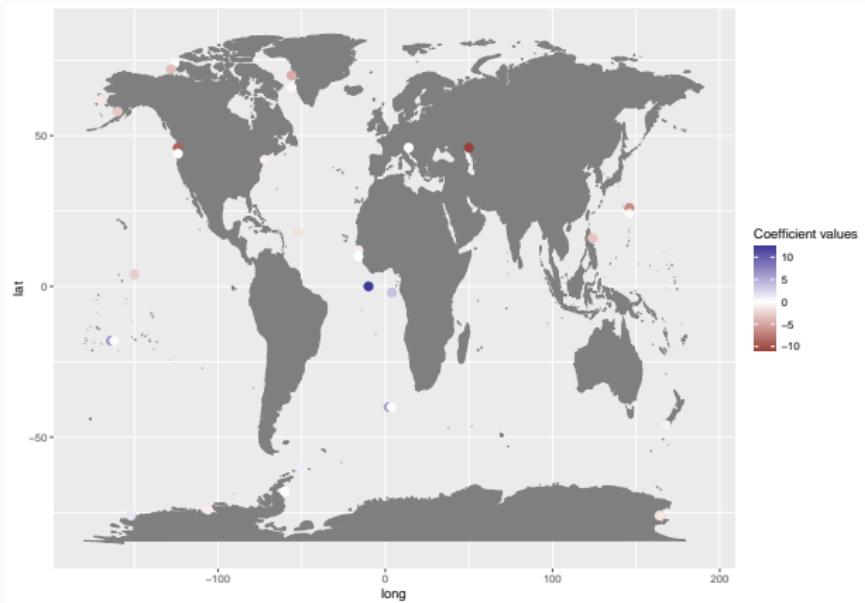
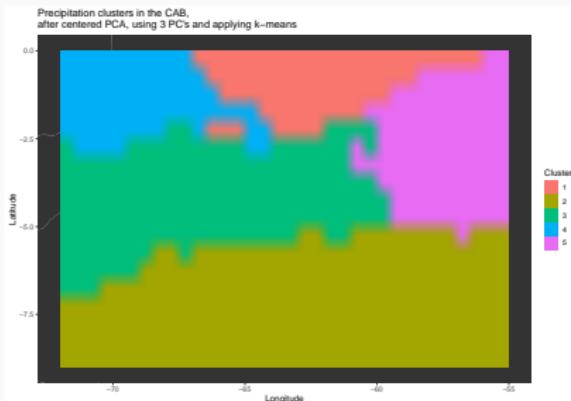


Figure 16: Coefficient plot of the full lasso model.

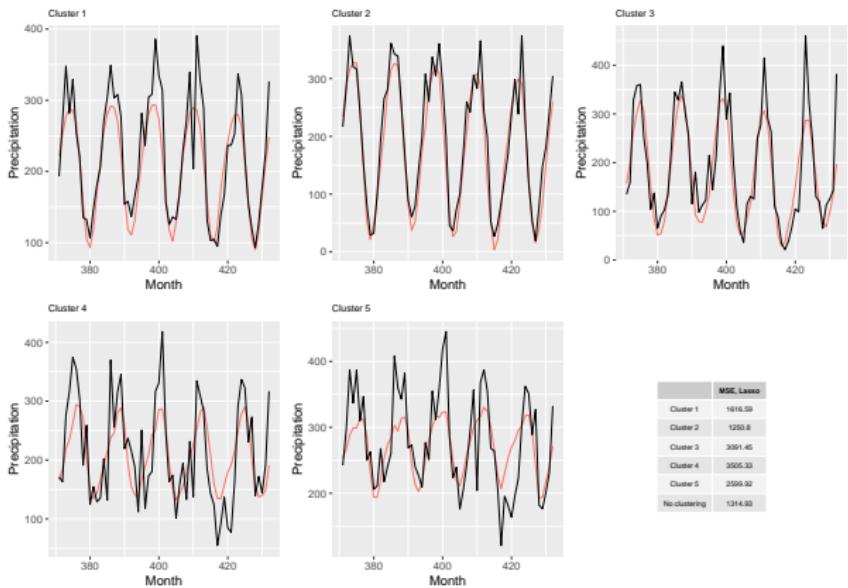
Lasso results all models

	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



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 too far away from hold-out time frame
- Our validation approach is a trade-off between efficient use of

6. The fused lasso

Definition of the fused lasso

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

$$\min_{\beta} \frac{1}{2} \sum_{i=1}^n (y_i - x_i^T \beta)^2 + \lambda \sum_{i,j \in E} |\beta_i - \beta_j| + \gamma \cdot \lambda \sum_{i=1}^p |\beta_i|, \quad (6)$$

- with x_i being the i th 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

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|D\beta\|_1, \quad (7)$$

- 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 $\text{rank}(X) = p$ (this is called the “signal approximator” case), the dual problem of (7) is then:

$$\hat{u} \in \arg \min_{u \in \mathbb{R}^\omega} \frac{1}{2} \|y - D^T u\|_{\frac{2}{2}} \text{ subject to } \|u\|_\infty \leq \lambda. \quad (8)$$

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

$$\hat{\beta} = y - D^T \hat{u}. \quad (9)$$

- 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

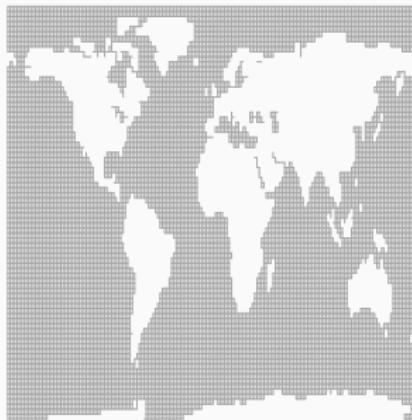


Figure 17: Graph of the SST and land areas used in fused lasso

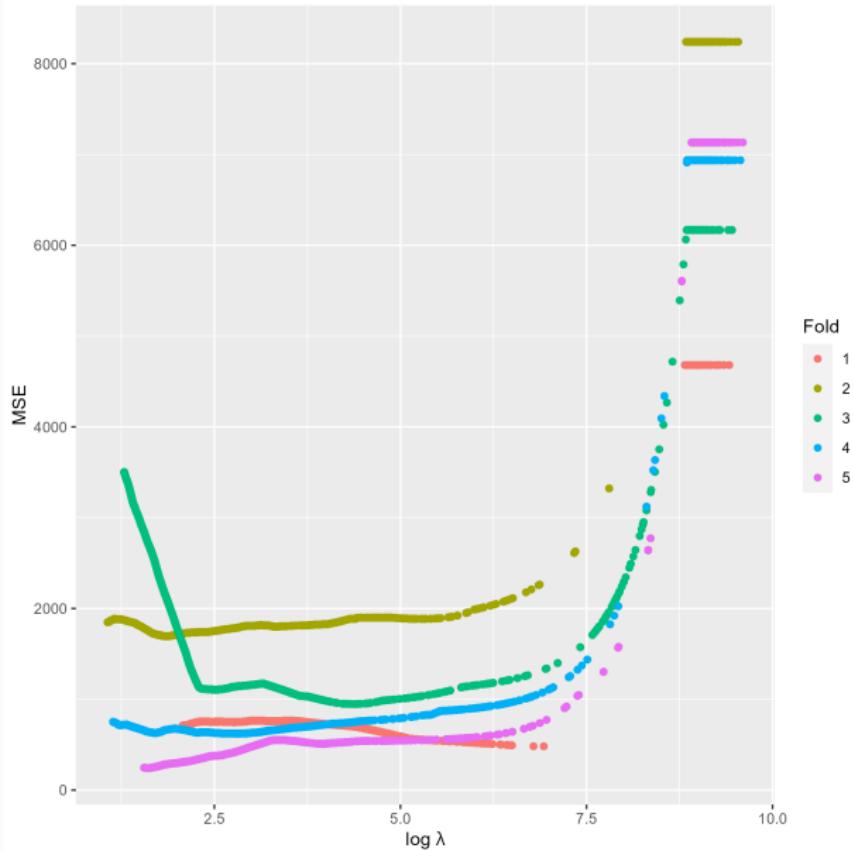
Graph structure and implications

- 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

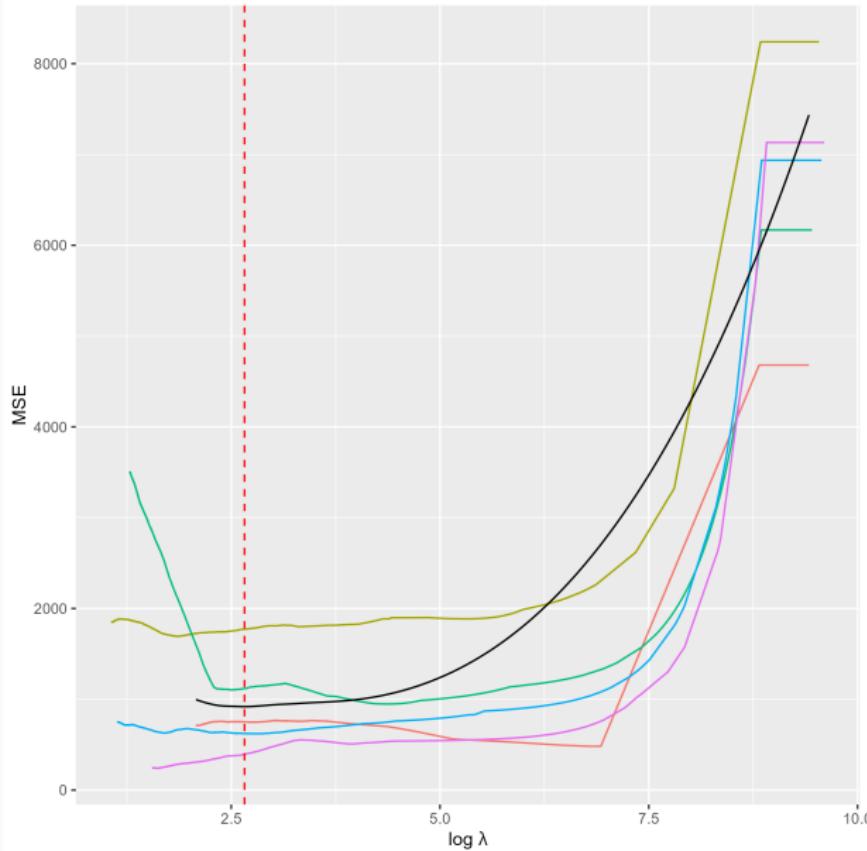
Fused lasso settings

- 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

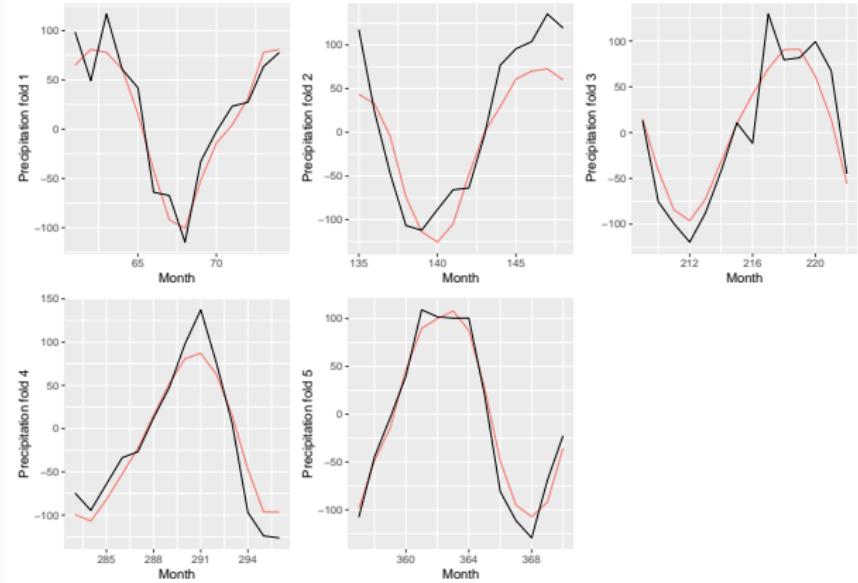


Figure 18: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

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

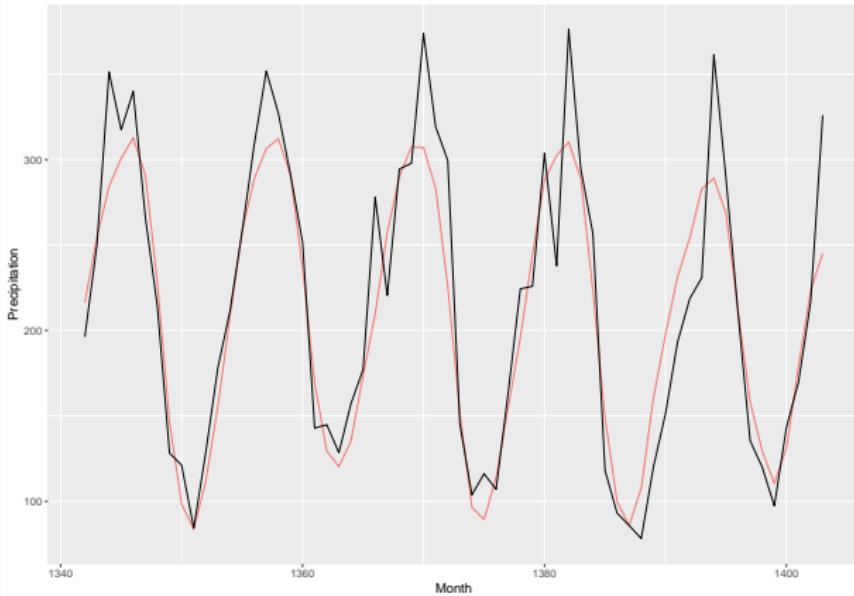
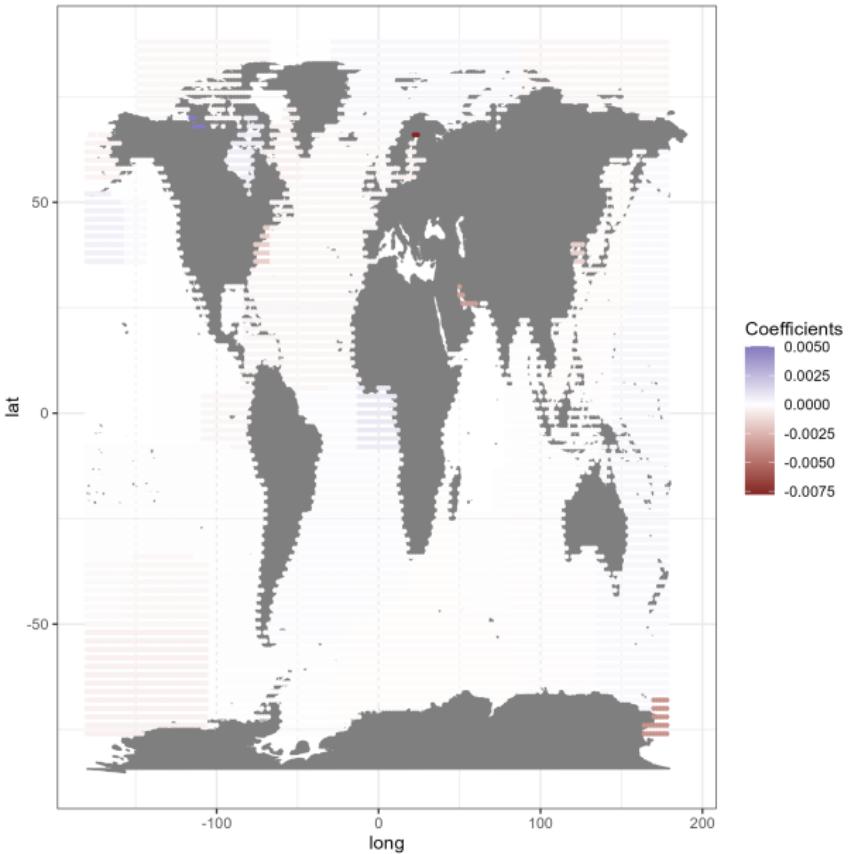
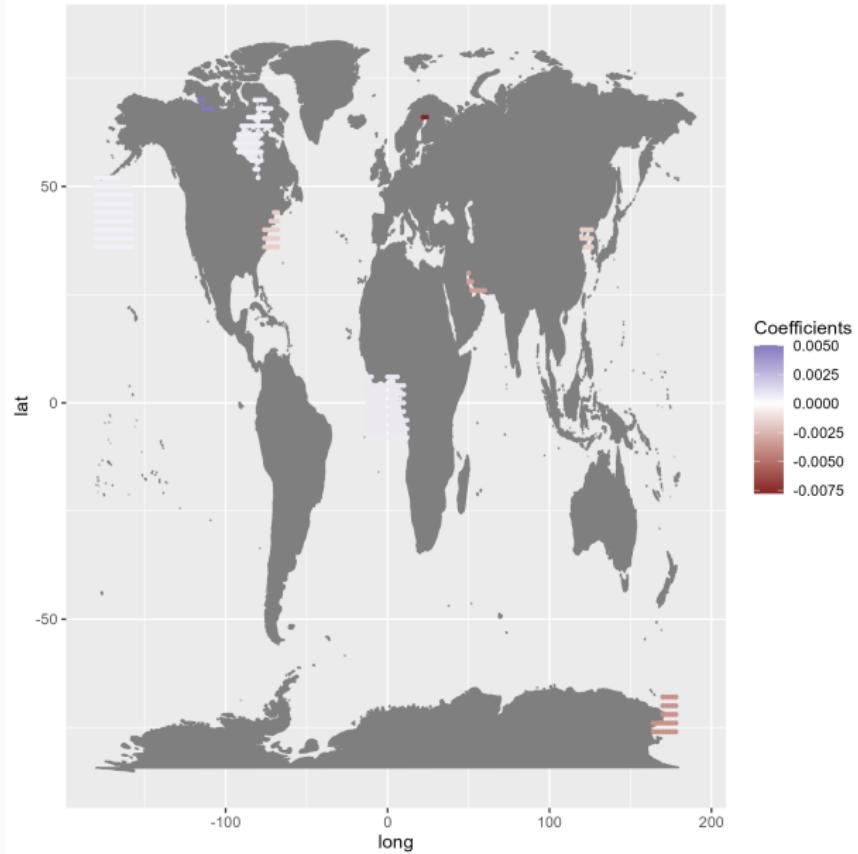


Figure 19: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

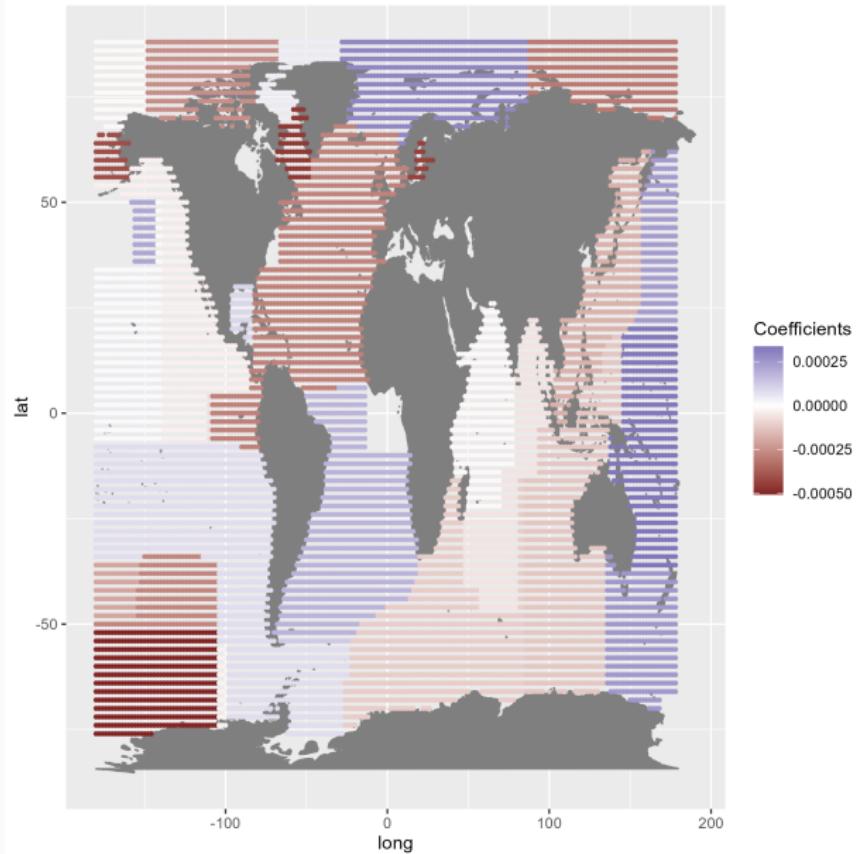
Coefficient plot



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

7. 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.

8. References

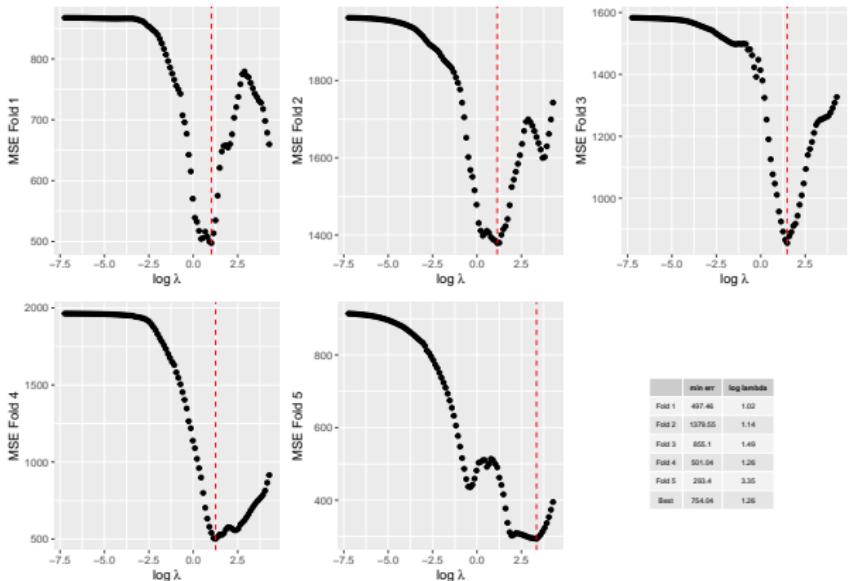
8. References

- 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.
- Tibshirani, RJ. 2011. "Regression Shrinkage and Selection via the Lasso."
- 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.
- Van der Kooij, Anita J. 2007. *Prediction Accuracy and Stability of Regression with Optimal Scaling Transformations*. Leiden University.

Appendix

Lasso on original SST

MSE in each fold (Lasso on original SST)



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

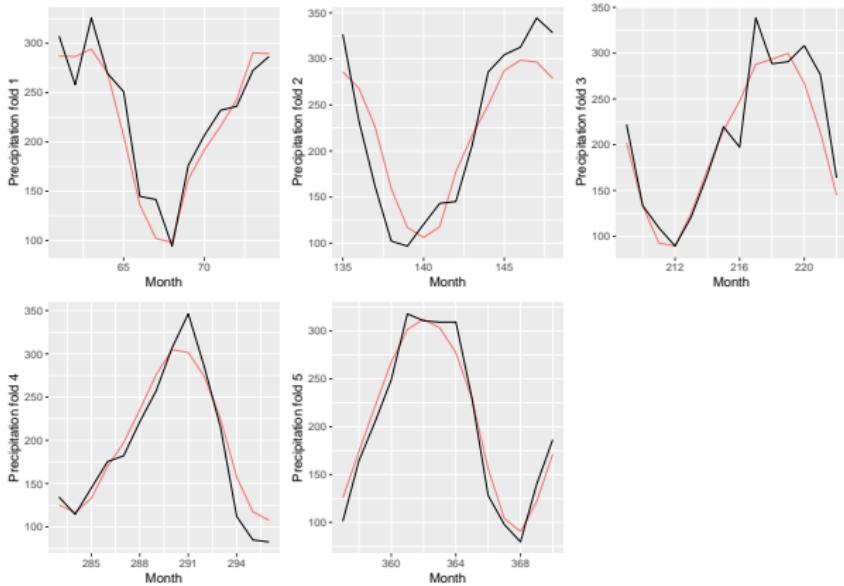


Figure 20: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set (Lasso on original SST)

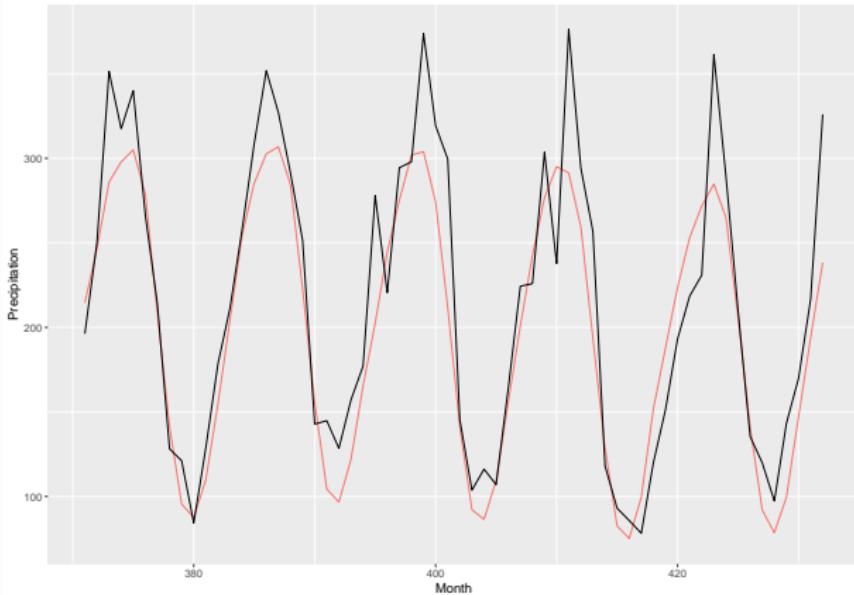


Figure 21: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

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

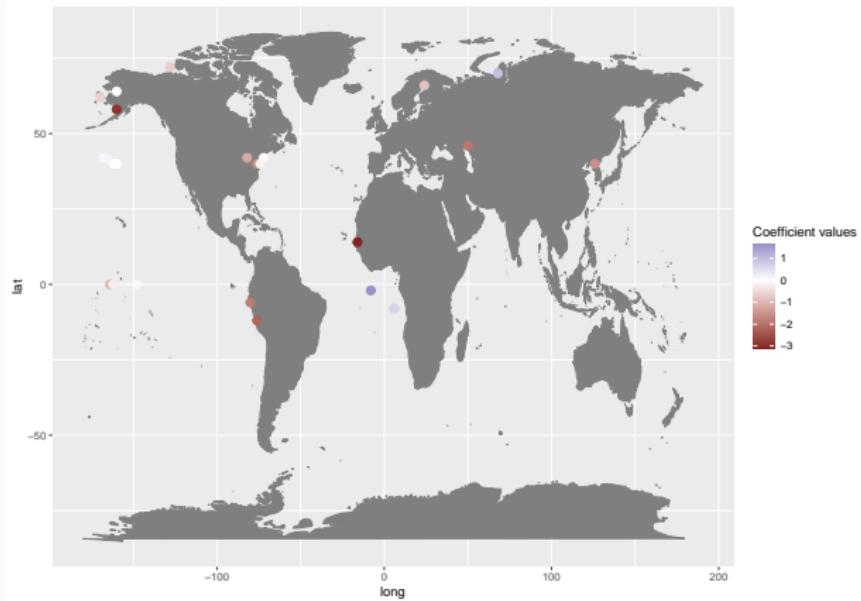
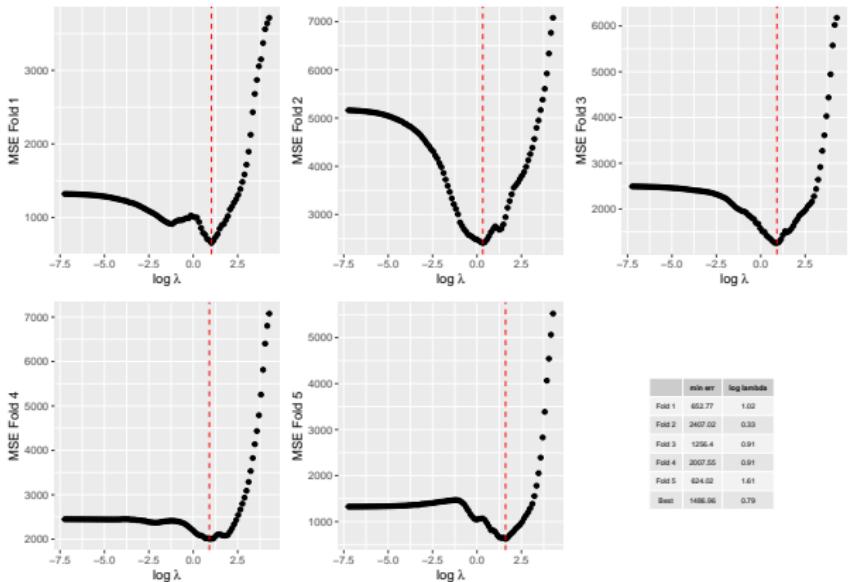


Figure 22: Coefficient plot of the full lasso model.

Lasso on differentiated SST

MSE in each fold (Lasso differentiated)



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

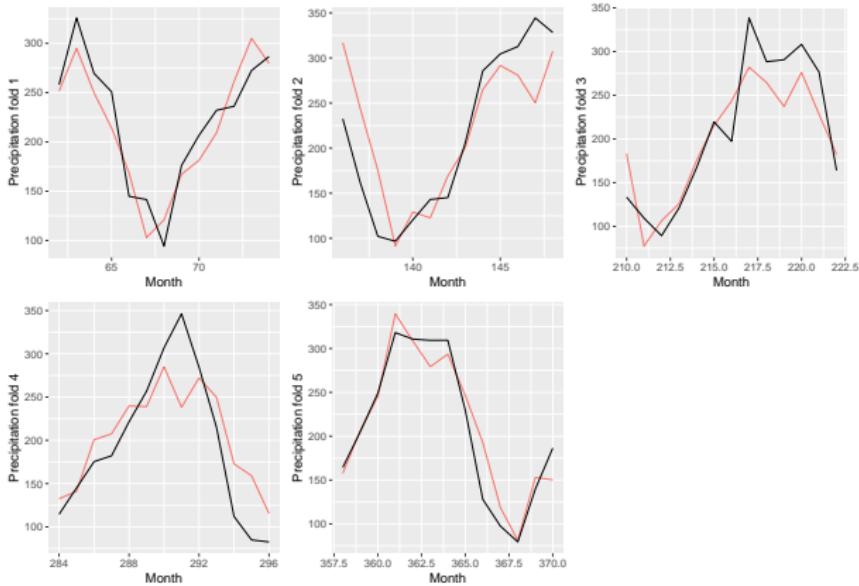


Figure 23: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set (Lasso differentiated)

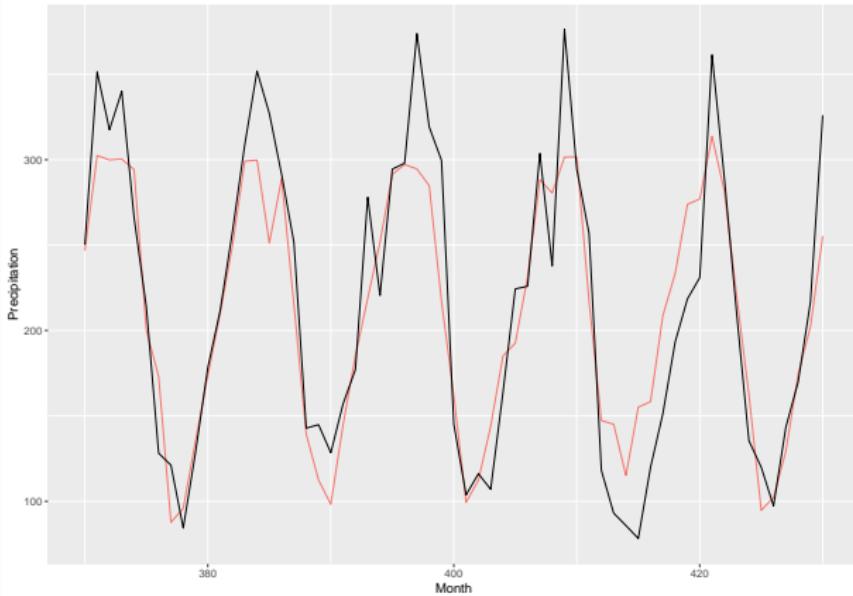


Figure 24: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

SST Regions chosen by the lasso (Lasso differentiated)

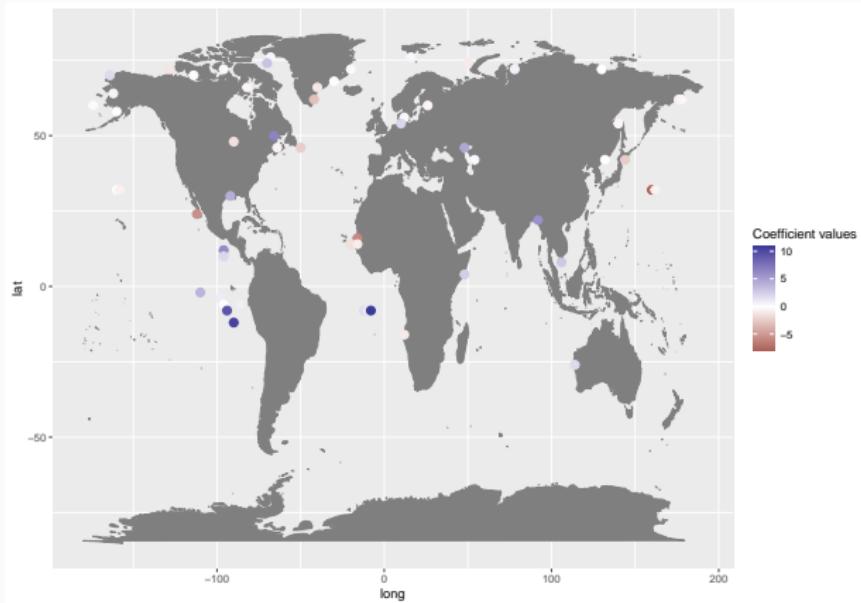


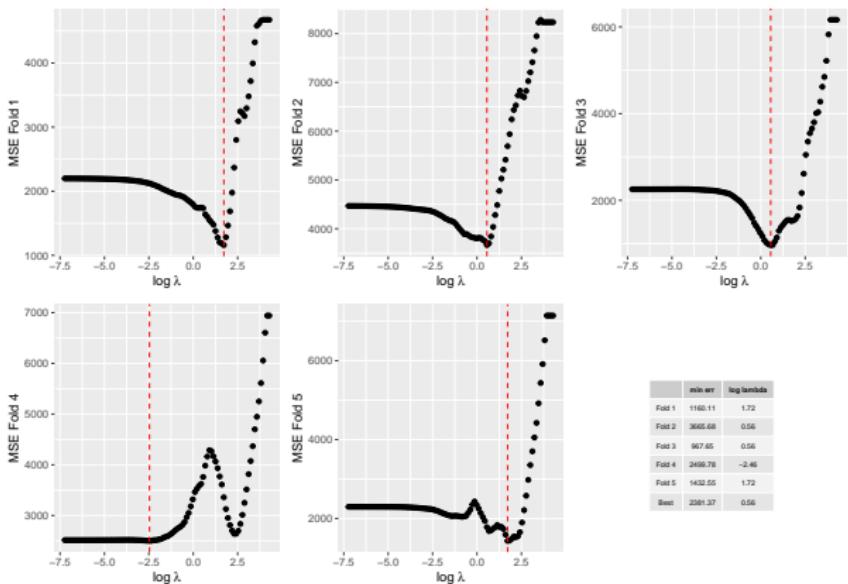
Figure 25: Coefficient plot of the full lasso model.

Fused evaluation (maybe explain this when showing results)

- 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)



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

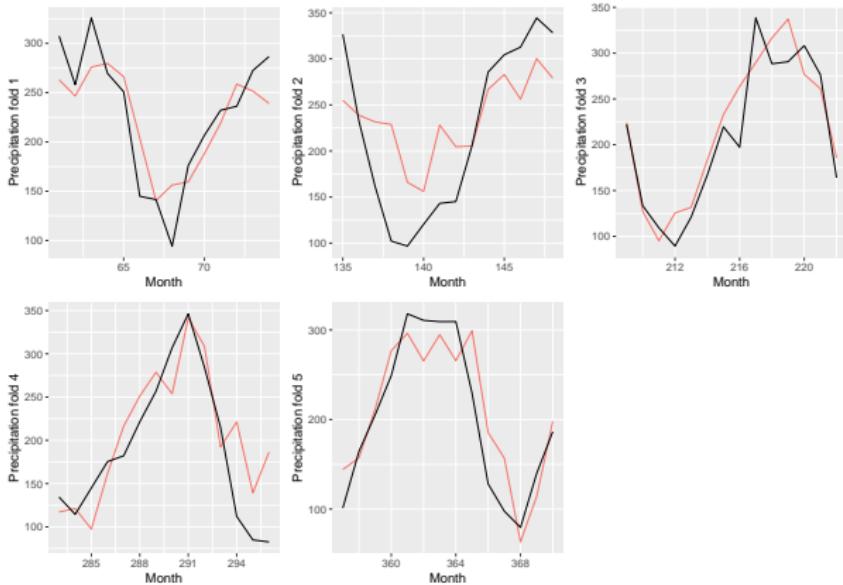


Figure 26: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set (Lasso de-seasonalized)

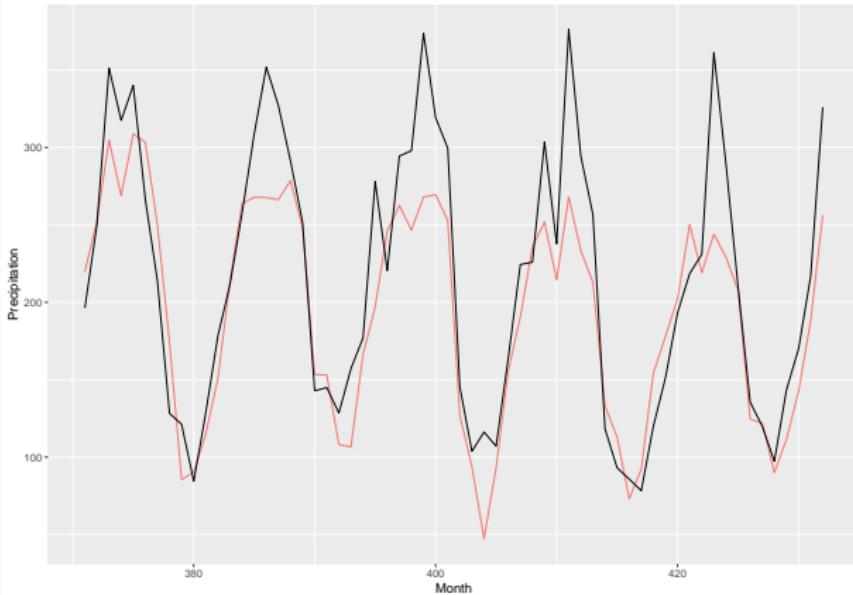


Figure 27: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

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

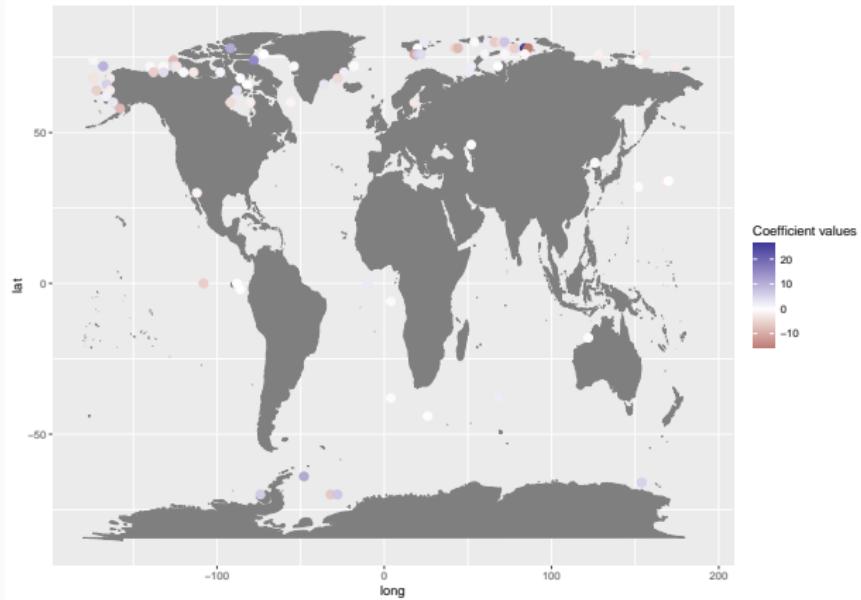
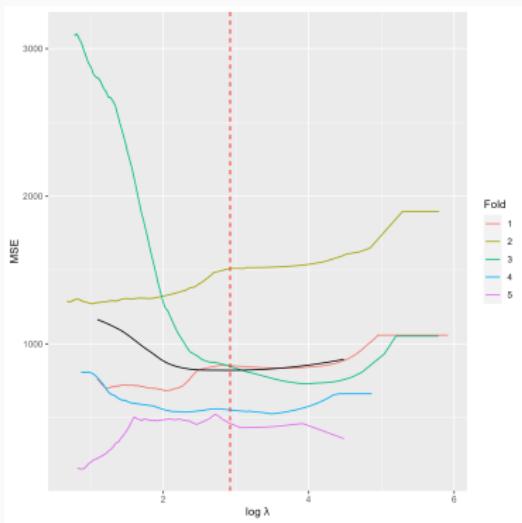


Figure 28: Coefficient plot of the full lasso model.

Fused lasso with sub-graphs

Error lines (Fused lasso with sub-graphs)



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

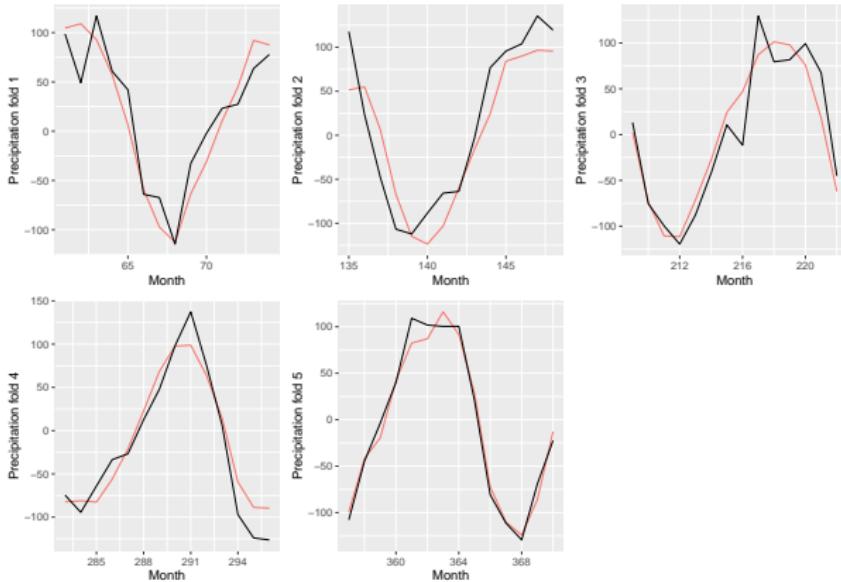


Figure 29: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

The predictions inside the folds are very similar to lasso without standardization (see 20), the same holds for the predictions from

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

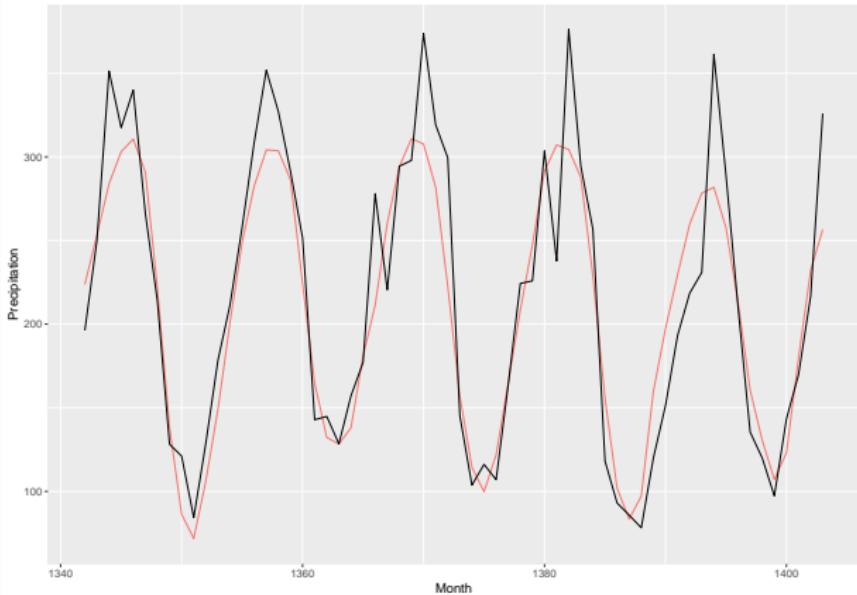
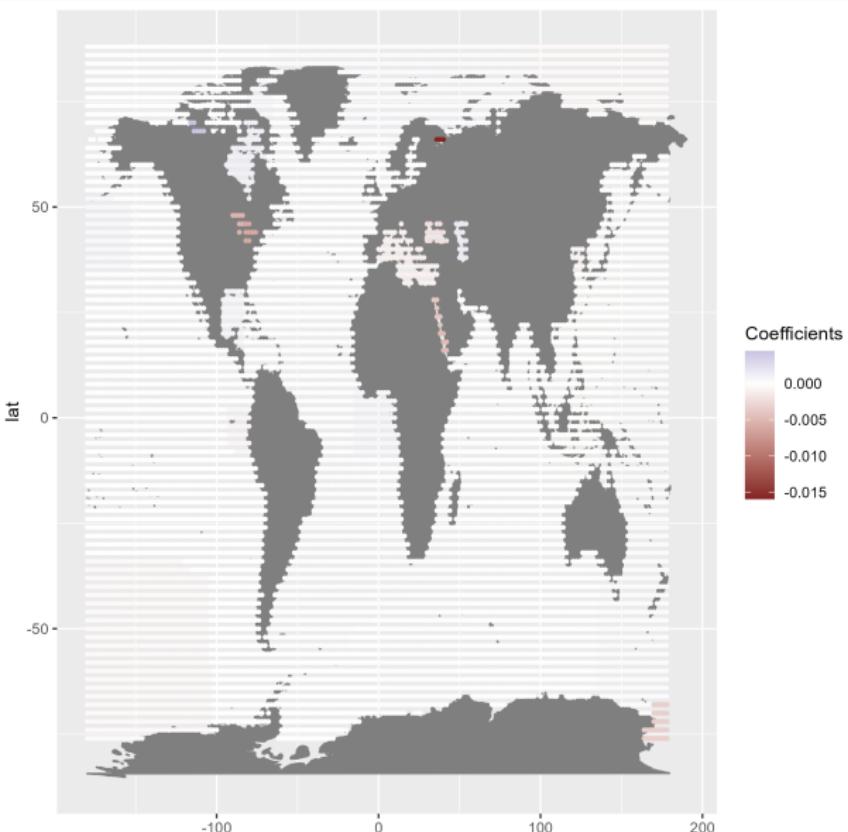


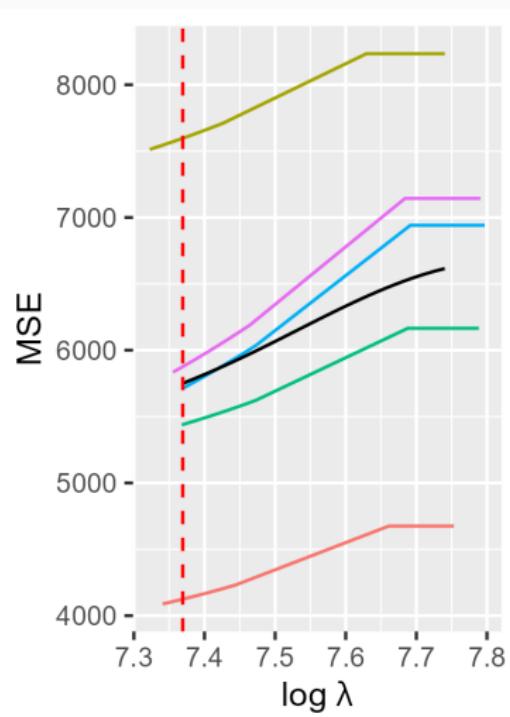
Figure 30: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

Coefficient plot of the final model (Fused lasso 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)

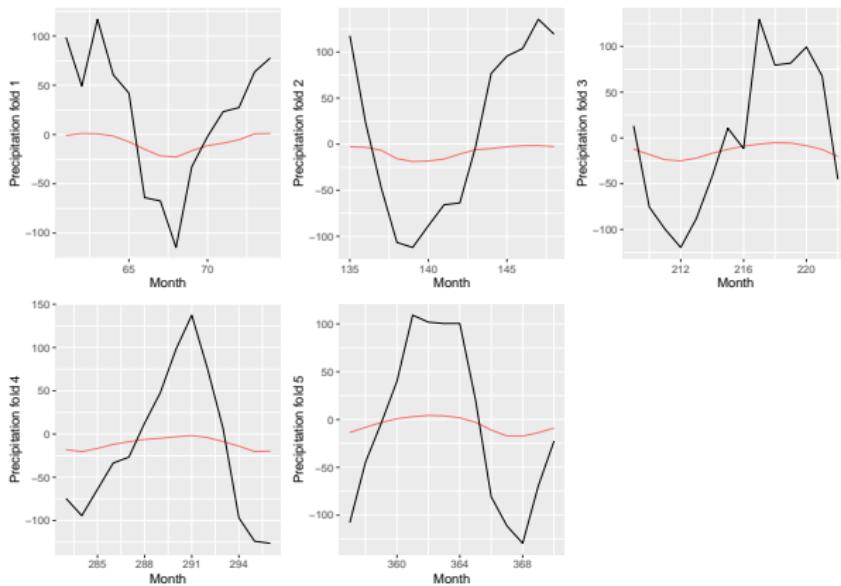


Figure 31: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

The predictions inside the folds are very similar to lasso without

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

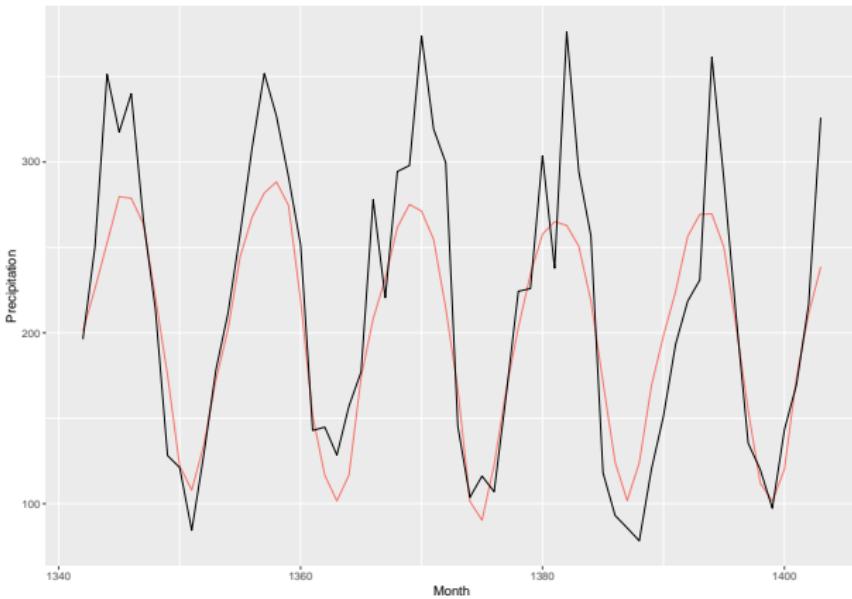
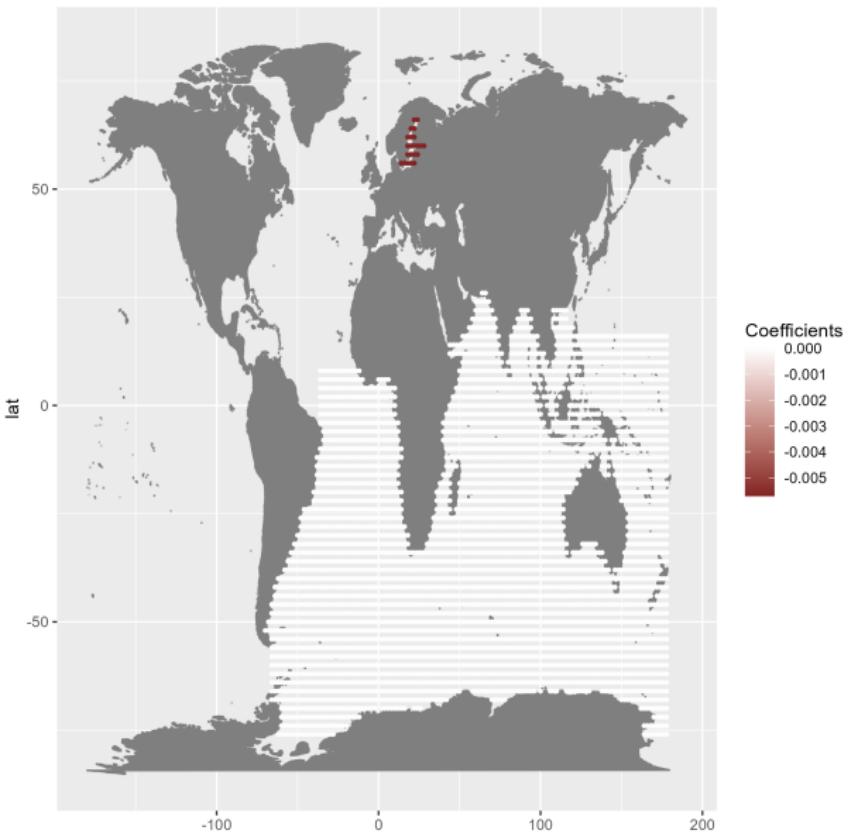


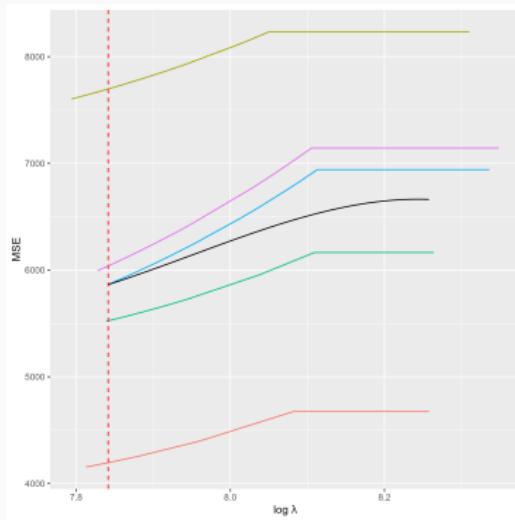
Figure 32: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

Coefficient plot of the final model (Fused lasso without sub-graphs, 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)

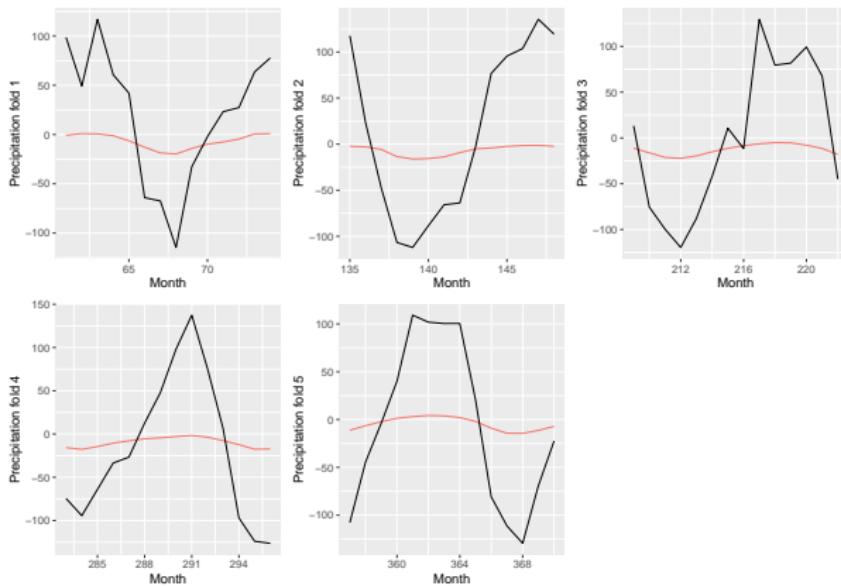


Figure 33: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

The predictions inside the folds are very similar to lasso without

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

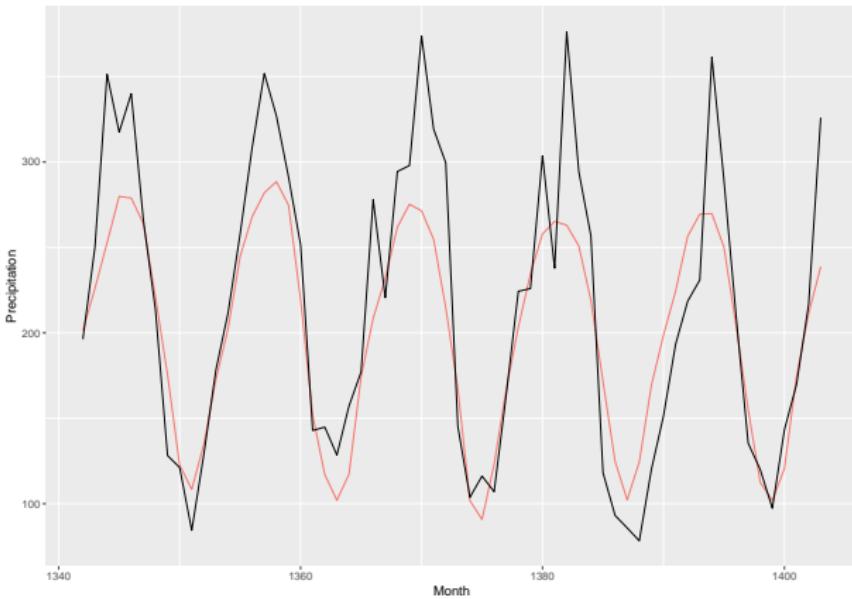
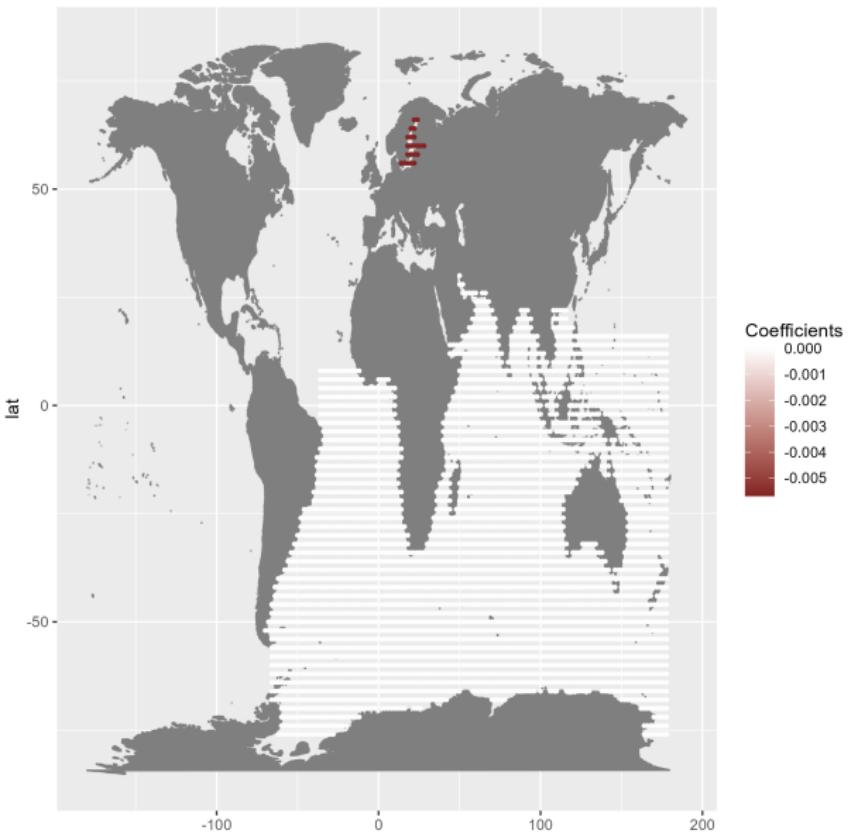


Figure 34: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

Coefficient plot of the final model (Fused lasso without sub-graphs, gamma 0.05)



**Fused evaluation (maybe explain
this when showing results)**

Fused evaluation (maybe explain this when showing results)

- 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