

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 crucial ecosystem for biodiversity, carbon storage, and moisture recycling.
- Hydrological extremes can affect this ecosystem, populations and biomass carbon impact tremendously.
- Severe Amazon drought in 2010 had total biomass carbon impact of 2.2 PgC , affected area 3 Mio km^2 (Lewis et al. 2011).
- Ciemer et al. (2020) established an early warning indicator for water deficits in the central Amazon basin (CAB).

Related work

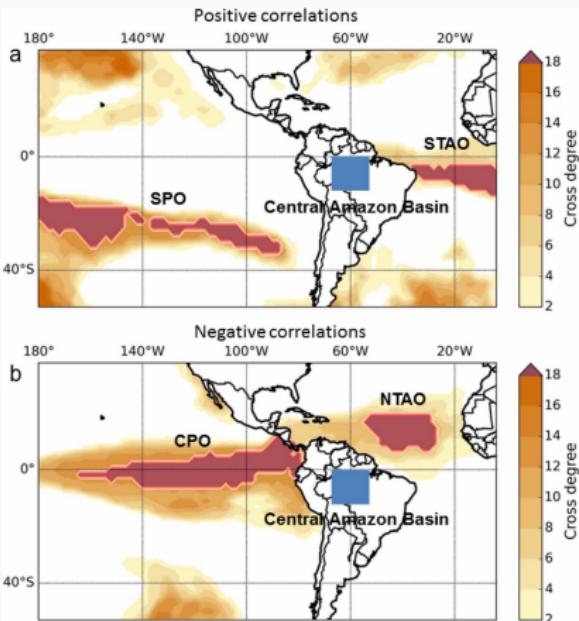


Figure 1: Cross degree between sea surface temperature and continental rainfall anomalies (Ciemer et al., 2020))

Our approach

- Inspect spatial and temporal characteristics in the raw data.
- Fit different models to predict precipitation directly from sea surface temperature (SST).
- Apply the lasso and fused lasso with different parameter settings (R. Tibshirani (1996) and R. Tibshirani et al. (2005), respectively).
- Evaluate the models with model validation techniques for time series.

2. Explorative analysis

The data

- Rain data from CHIRPS in *mm/month*(Climate Hazards Group InfraRed Precipitation with Station data, Funk et al. (2015)).
- CHIRPS is derived from in-situ and satellite data.
- SST data from ERSST (Extended Reconstructed Sea Surface Temperature, Huang et al. (2017)).
- ERSST is a 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), over the period 1981 until 2016 (36 years).

Explorative analysis rain



Figure 2: Location of the area under study. The central amazon basin (CAB) spanning across 0,-10 latitude and -70,-55 longitude.

Precipitation, mean, and SD

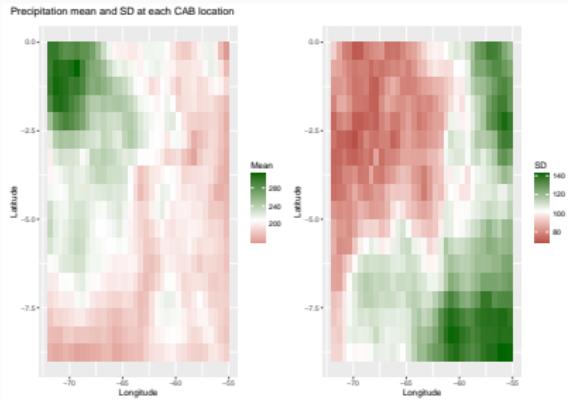


Figure 3: Mean and standard deviation at each location. The standard deviation was computed over the whole time period. The white lines on the legend at each side of the plots indicate the mean of the respective quantity.

Glyph lots

- The mean and standard deviation in the CAB don't show the development over time.
- We can display time series with spatial structure with so-called *glyph-plots*.
- Here, a glyph is one time series of precipitation shown at its location in the CAB.
- To see the trajectory of the time series more clearly, we show them after removing seasonality and scale them as:

$$x_{rescaled} = \frac{x - \min(x)}{\max(x) - \min(x)}. \quad (1)$$

Precipitation glyph plots

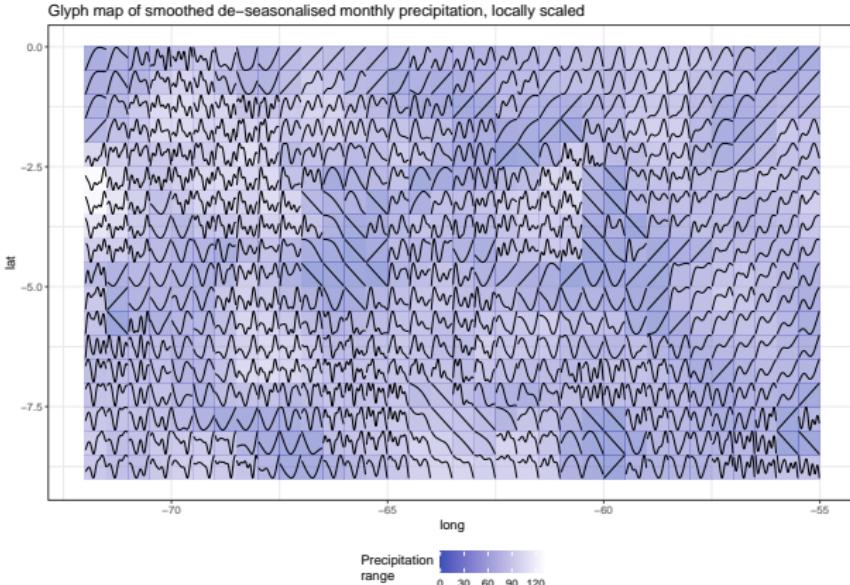


Figure 4: Glyph map of de-seasonalised and smoothed precipitation. The time series are scaled locally; ranges are not the same in all cells. The different ranges are given in color shades, where lighter shading indicates a larger range and darker shades smaller ranges.

Explorative analysis SST

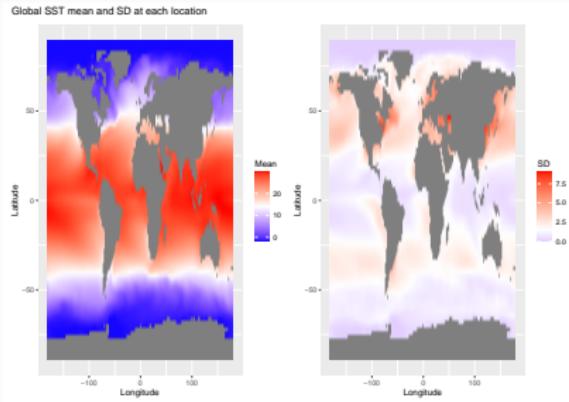


Figure 5: Mean and SD SST on the global map. The white lines on each legend at the side of the plots indicate the mean of the respective quantity.

3. Correlation analysis

Correlation analysis

- Here we get an overview of 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 (spurious correlations).

Correlation plot original SST

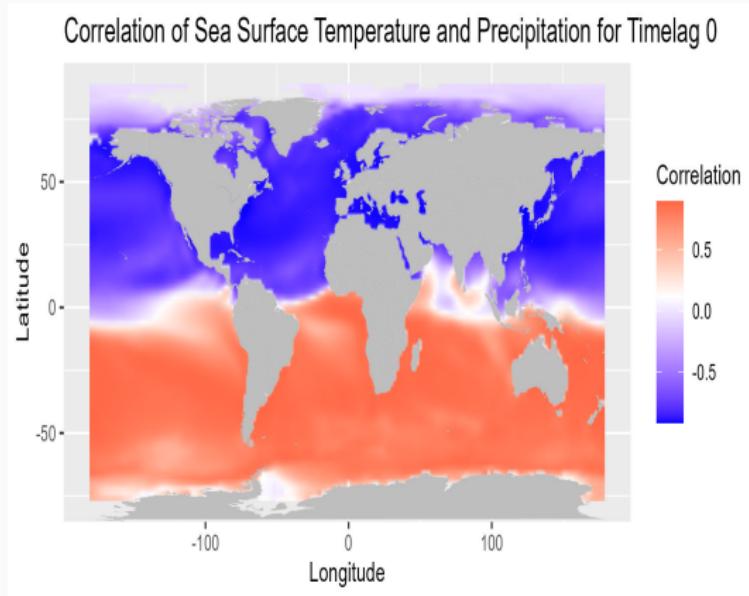


Figure 6: Correlation plot between SST and mean precipitation in the CAB for timelag 0.

Correlation plot original SST

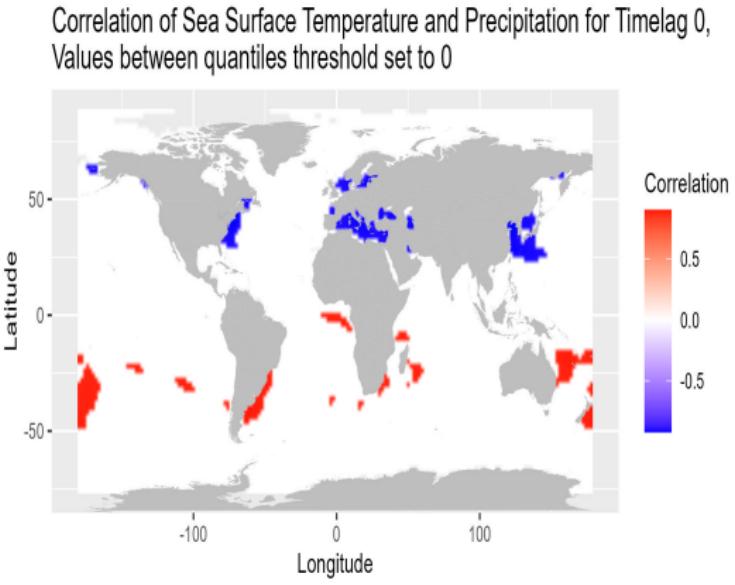


Figure 7: Correlation plot between SST and mean precipitation in the CAB for time-lag 0. Values between the 97.5% and 2.5% quantiles are set to 0.

Correlation plot de-seasonalized SST

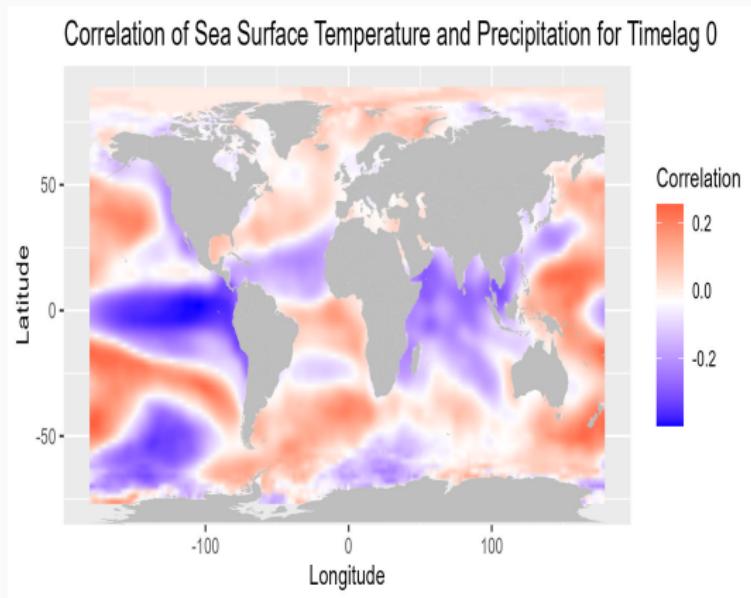


Figure 8: Correlation plot between de-seasonalized SST and de-seasonalized mean precipitation in the CAB for timelag 0.

Correlation plot de-seasonalized SST

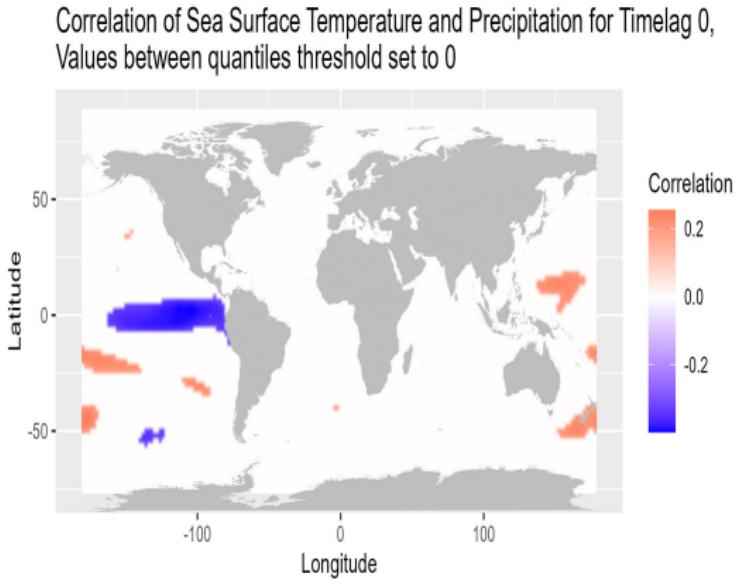


Figure 9: Correlation plot between de-seasonalized SST and de-seasonalized mean precipitation in the CAB for timelag 0. Values between the 97.5% and 2.5% quantiles are set to 0.

4. Clustering precipitation

Clustering motivation

- Our explorative analysis has shown spatial and temporal differences in the precipitation data.
- We explored this further using k -means clustering:
 - found optimal k via PCA and gap statistic
 - applied k -means to the three first PC of the precipitation data
- Ultimate goal is to improve predictions by applying a model to each cluster separately.

Clustering results

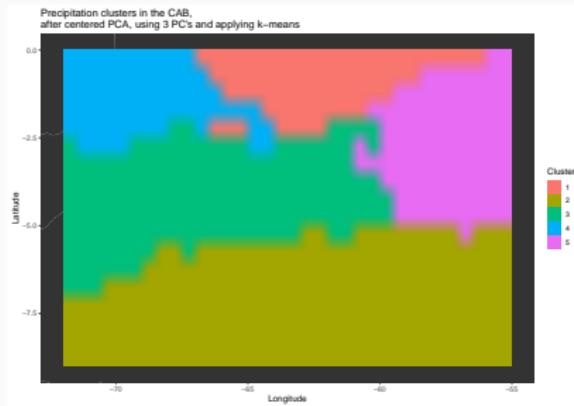


Figure 10: Spatial organization of the found clusters in the CAB. We applied a centered PCA on the data and used three principal components before applying the k-means algorithm

Clustering results

- We find 5 clusters of different sizes.
- The found clusters are almost entirely spatially coherent although we did not include any spatial dependencies in the clustering.
- The exception is the “island” of cluster 1 (orange) inside cluster 4 (blue) and on 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 (R. Tibshirani 1996}).
- We consider the lasso regression problem:

$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N l(y_i, \beta_0 + \beta^T x_i) + \lambda \|\beta\|_1 \quad (2)$$

- $\lambda \|\beta\|_1$ introduces sparsity in the coefficient vector.
- The problem is solved using coordinate descent (Van der Kooij 2007).
- We use the glmnet package (Friedman, Hastie, and Tibshirani 2010).

Model validation

- To find the best λ , we can use k-fold cross validation to achieve a Bias-Variance trade-off (Hastie et al. 2009).
- We introduce bias by using fewer parameters ($\lambda > 0$) but decrease the prediction variance by doing so.
- Classic cross-validation uses the data efficiently, but we want to avoid predicting past values learned from future data.

Forward validation

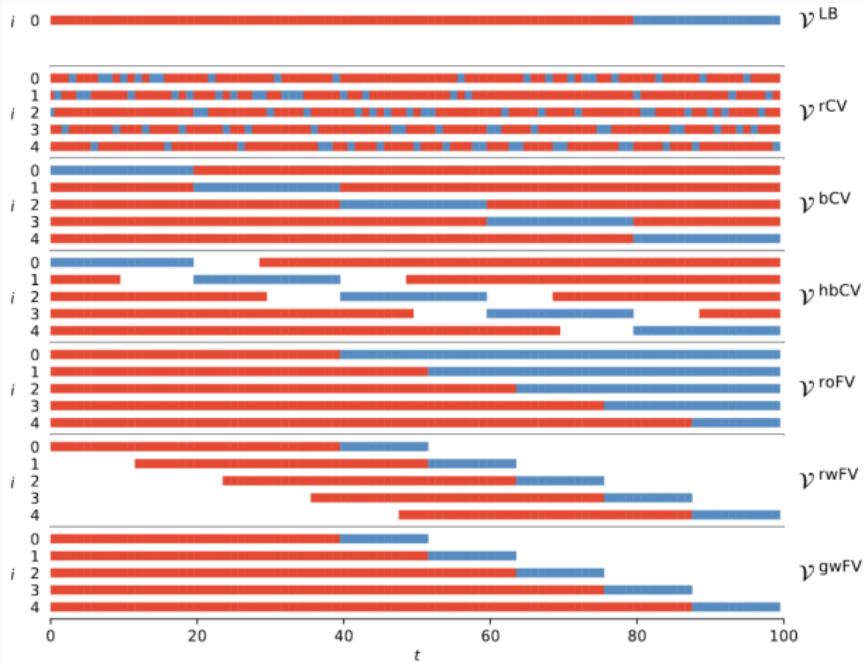


Figure 11: Different schemes of validation approaches for time series data (Schnaubelt 2019).

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 five folds.
- We also use some part of the data as a hold-out test set that is not used in forward validation.
- The forward validation set consists of the first 360 months, and the hold-out test contains the subsequent 62 months (months 371 to 432).
- Each fold has 60 months of training and 14 months of the test.

Forward validation

- We compute a λ -vector of length l for the complete model validation set.
- For each fold, we fit l models with this λ -vector.
- On each FV test set, we compute l prediction errors.
- Since we have five folds, we have five prediction errors for each value of the λ -vector.
- Choose λ_{\min} as the λ that minimizes average MSE over all folds.
- Fit model on complete FV data with λ_{\min} and compute MSE on the hold-out test set.

Lasso settings and results

- We compare different settings for the lasso:
 - Lasso
 - Standardized SST
 - De-seasonalized SST
 - Differentiated SST
 - Lasso on clusters found by k -means
- If we standardize or de-seasonalize we use the information from model fitting to transform the test data.
- We do this to avoid *information leakage*.
- Lasso with standardized SST gave the best results.
- Clustering only improved on one cluster.
- Here, we show the results from the lasso with standardized SST and the final results from the lasso on clusters.

Lasso - MSE on FV test sets

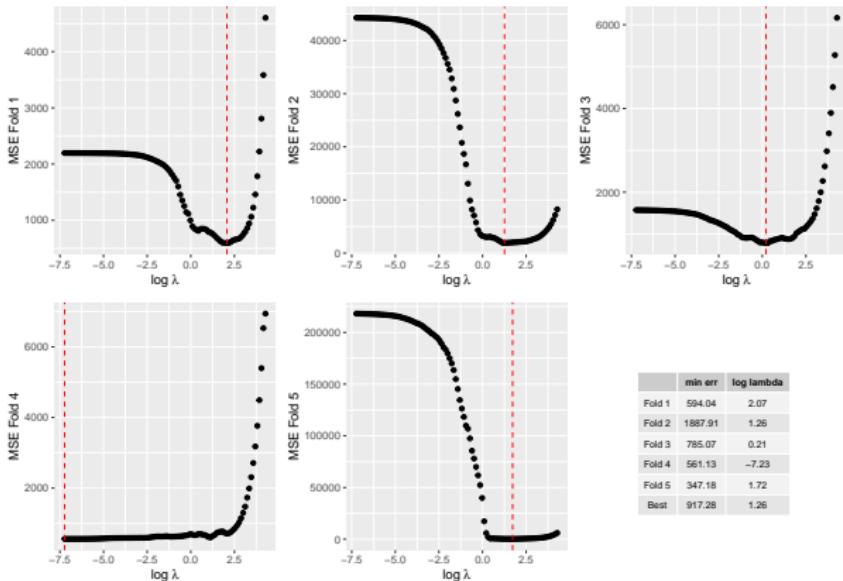


Figure 12: MSE of the CV for the different lambda values on the log scale. The red dotted line shows the lambda for which minimum MSE was obtained.

Lasso - Predictions on FV test sets

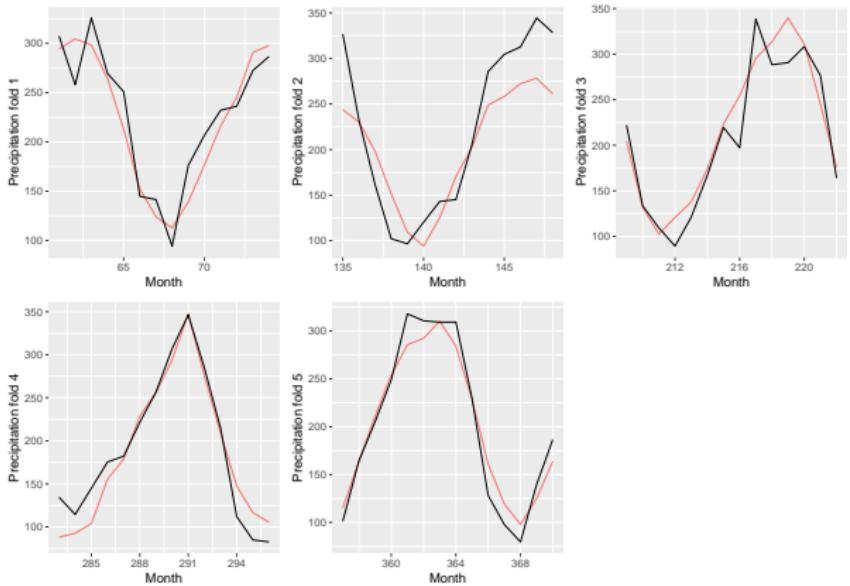


Figure 13: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Lasso - Predictions on the hold-out test set

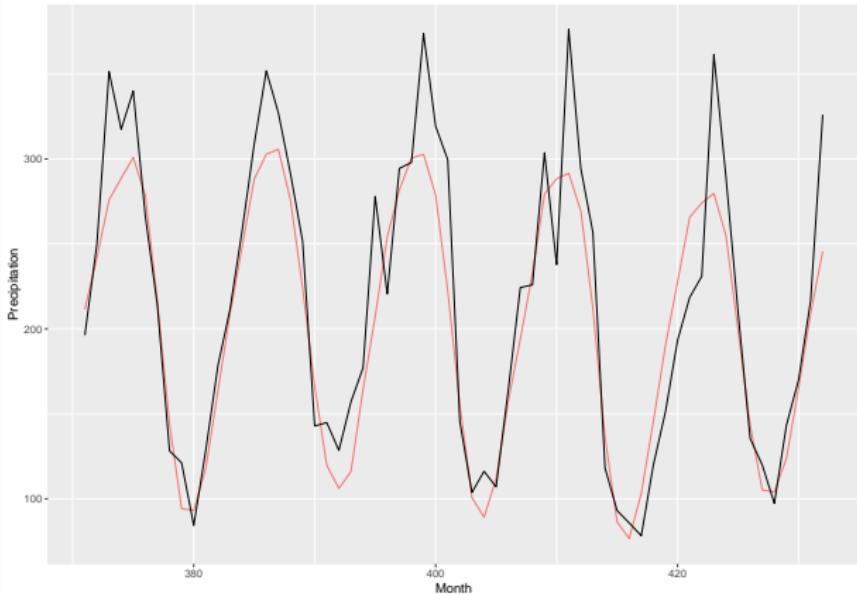


Figure 14: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE.

Lasso - Predictions on the hold-out test set

- The final lasso model predicts seasonal cycles without major peaks or drops.
- It constantly fails to predict the spikes in precipitation.
- Low values are predicted better than high values.
- The predicted amplitudes get smaller over time.

Lasso - SST Regions, final model

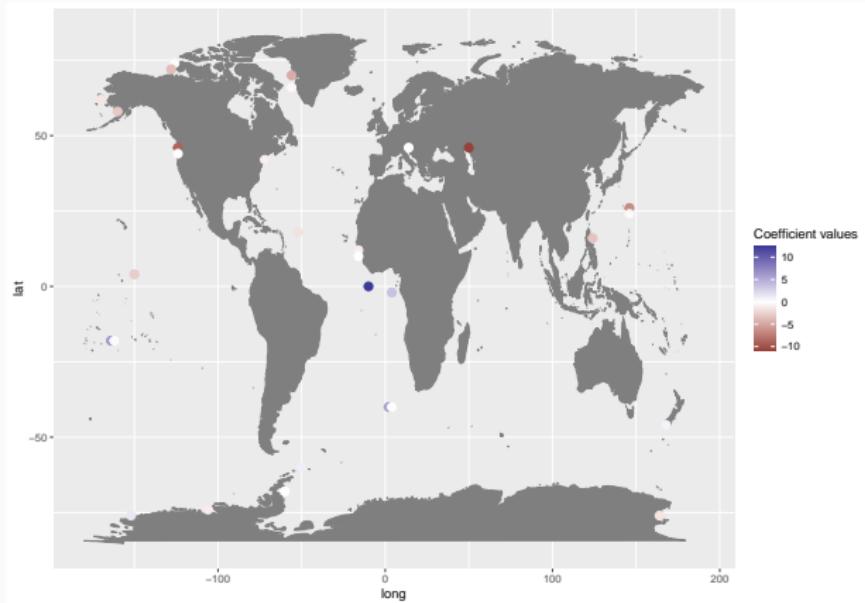


Figure 15: Coefficient plot of the full lasso model.

Lasso - SST Regions, final model

- The lasso chooses coefficient values as single “points” that are generally not close to each other.
- The lasso uses no information about the coefficients’ distances to each other.
- Some areas with large coefficient values also showed high values in the correlation analysis.
- From an area with high correlation, the lasso only chooses one “point” and discards the others.

Lasso - All model results

Table 1: Table of the MSE results for the different lasso models evaluated. The λ refers to the value that was found in forward validation and used to fit the final model.

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

Lasso - Results on clusters

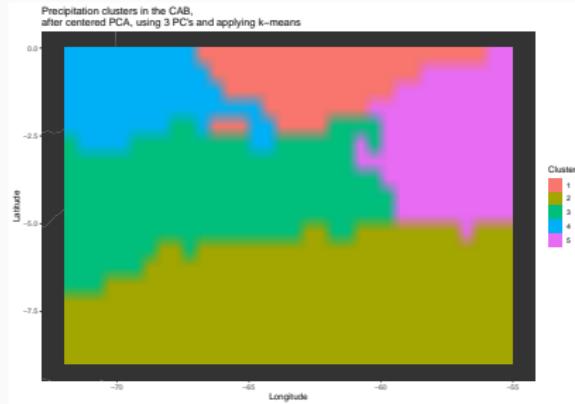


Figure 16: Spatial organization of the found clusters in the CAB. We applied a centered PCA on the data and used three principal components before applying the k-means algorithm.

Lasso - Results on clusters

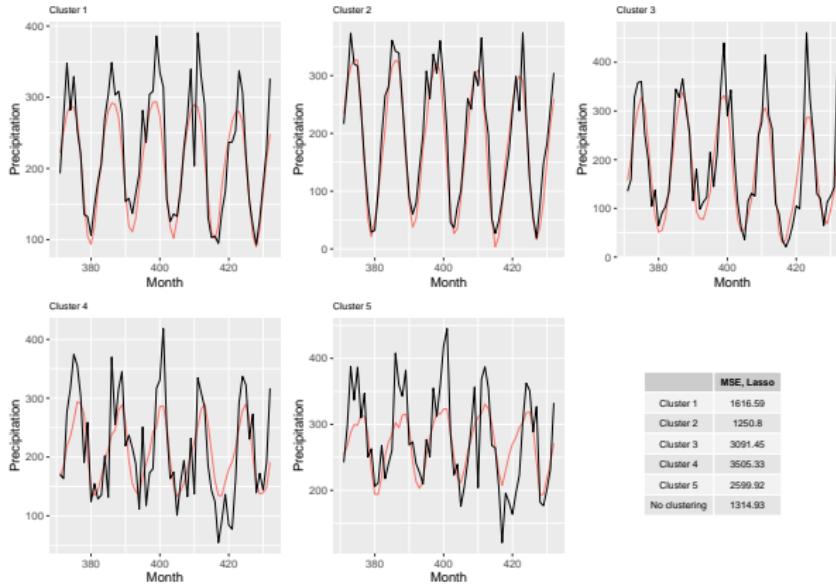


Figure 17: Precipitation prediction and target values in the validation set for the 5 clusters found with k -means. Predictions are in red and target values in black. The model was fitted on the full CV data with the lambda value that minimized the average MSE.

Summary lasso results

- We compared different settings for the lasso
 - Lasso
 - Standardized SST
 - De-seasonalized SST
 - Differentiated SST
 - Lasso on clusters found by k -means
- Lasso with standardized SST worked best.
- It can predict general seasonality but still fails to predict peaks in precipitation.
- Clustering the CAB improves predictions only on one cluster, but on this cluster, peaks can be predicted better than in the original model.

Summary lasso results

- We de-seasonalized SST to remove spurious correlations:
 - Used seasonality from train to de-seasonalize test data.
 - Can be a problem if seasonality differs from train to test, didn't improve results.
- To remove changing means and trends over time altogether, we removed non-stationarity.
 - Non-stationarity can be removed by differentiation, and we do not use information from training during testing.
 - But also loose information of the time series and did not improve either.

Summary lasso results

- The standard deviations differ significantly across the SST, and some of the locations are not really “sea” areas.
- Temperature changes can be more meaningful in some SST regions than in others.
- Standardizing then makes the comparison easier.
- This might be an explanation for why standardization improved our results here.

6. The fused lasso

Definition of the fused lasso

- The Fused lasso, “fuses” predictors together, and chooses connected regions instead of single “points” (R. Tibshirani et al. 2005).

$$\min_{\beta} \frac{1}{2} \sum_{i=1}^n (y_i - x_i^T \beta)^2 + \lambda \sum_{i,j \in E} |\beta_i - \beta_j| + \gamma \cdot \lambda \sum_{i=1}^p |\beta_i| \quad (3)$$

- with x_i being the i th row of the predictor matrix and E is the edge set of an underlying graph.
- The second term $\lambda \sum_{i,j \in E} |\beta_i - \beta_j|$ penalizes differences in connected coefficients.
- The third term $\gamma \cdot \lambda \sum_{i=1}^p |\beta_i|$, controls the sparsity of the coefficients, with higher values placing more priority on sparsity.

Fused lasso optimization

- Let's consider the problem in the notation of the generalized lasso problem (R. J. Tibshirani and Taylor 2011).

$$\hat{\beta} = \min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|D\beta\|_1 \quad (4)$$

- 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.
- In our case D is the edge set of a given graph, where m refer to the edges and p to the nodes.
- The problem is solved by the dual path algorithm, implemented in the *genlasso* package (Taylor B. Arnold and Tibshirani 2016).

Fused lasso graph structure

- We created a grid and deleted all nodes that were land regions, but this induced sub-graphs.

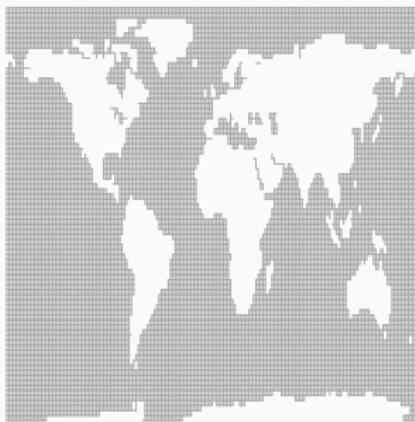


Figure 18: Graph of the SST and land areas used in the fused lasso.

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 the best results

Fused lasso evaluation

- Generally same setting as for lasso, five folds with train and test, choose λ_{\min} , refit with λ_{\min} , and get MSE on the hold-out test set.
- But, for the fused lasso, we can not define the λ vector beforehand.
- λ -path is found by the 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.

Fused lasso - MSE on FV test sets

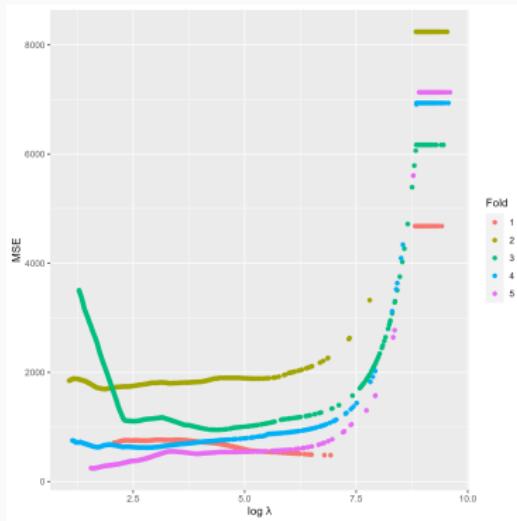


Figure 19: Error points in the fused lasso forward validation. The model uses a graph without sub-graphs. Each color represents one fold.

Fused lasso - MSE on FV test sets

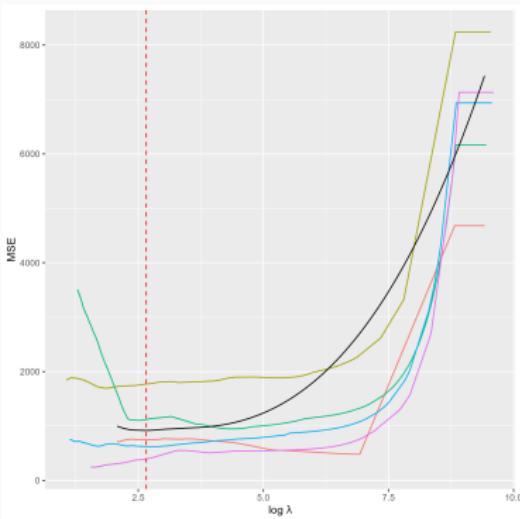


Figure 20: Error lines in the fused lasso forward validation. The model uses a graph without sub-graphs. Each line represents one fold, and the black line is the mean for the common interval of the smoothed error lines. The red dashed line shows the minimum of the mean.

Fused lasso - Predictions on FV test sets

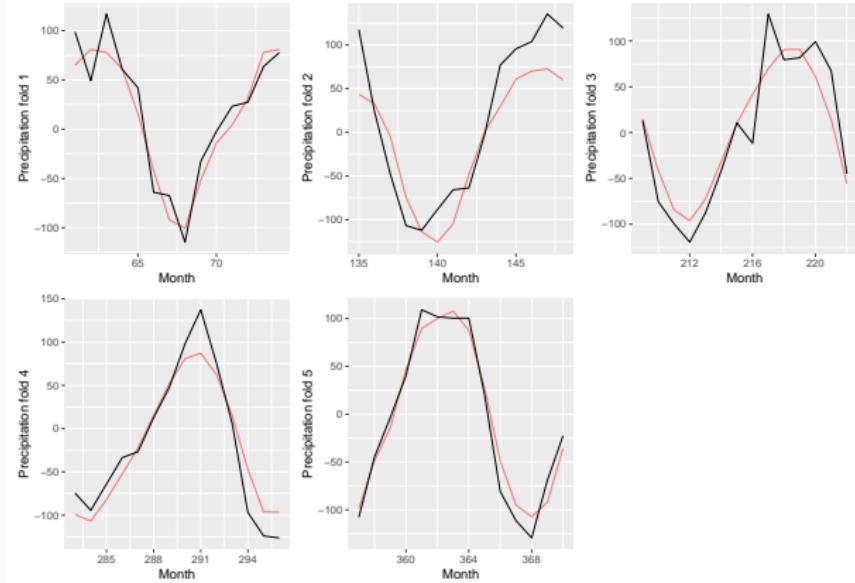


Figure 21: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Fused lasso - Predictions on FV test sets

- The predictions inside the folds are very similar to lasso without standardization.
- The same holds for the predictions from the full model.
- The MSE improves here and gives the best results overall.

Fused lasso - Predictions on hold-out test set

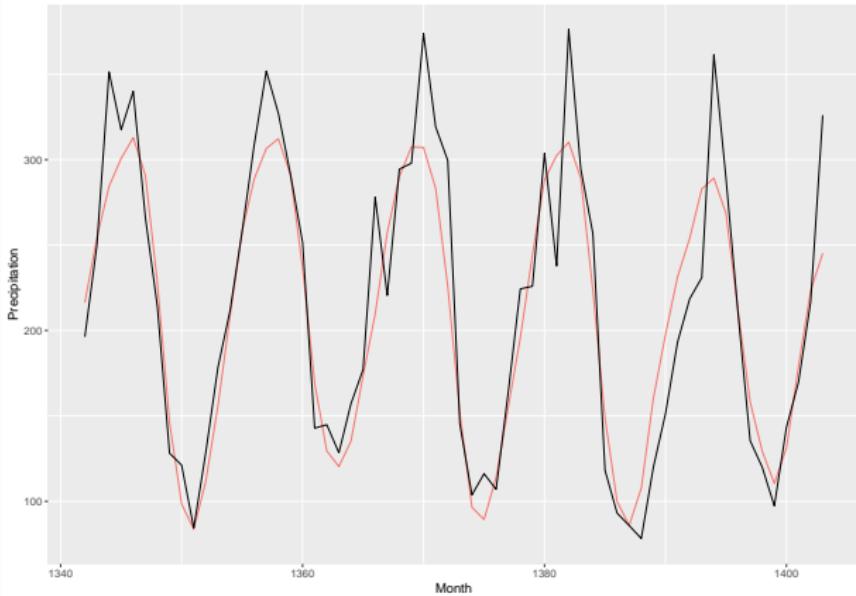


Figure 22: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE.

Fused lasso - SST regions, final model

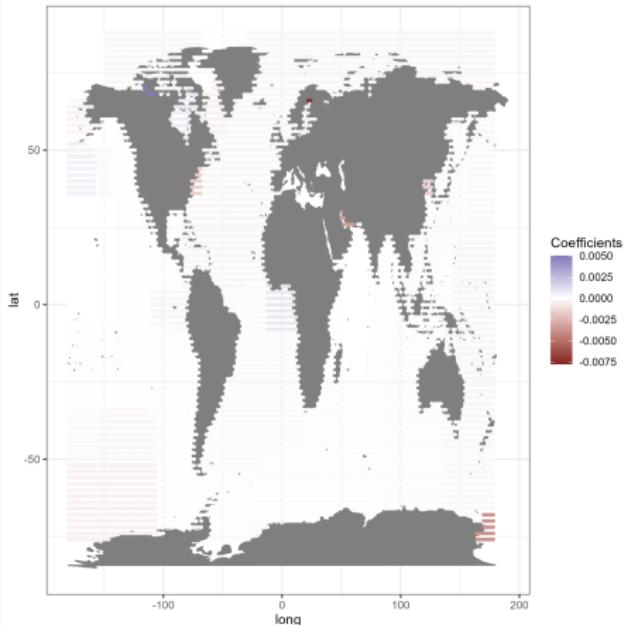


Figure 23: Coefficient plot of the final fused lasso model. Positive coefficient values are given in blue, negative values in red.

Fused lasso - SST regions, final model

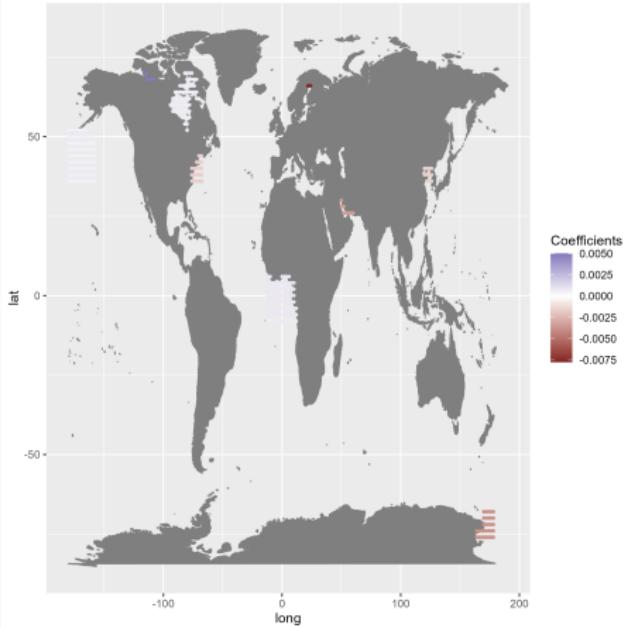


Figure 24: Coefficient plot of the final fused lasso model. Positive coefficient values are given in blue, negative values in red. Only values outside the 0.975 and 0.025 quantile range are shown.

Fused lasso - SST regions, final model

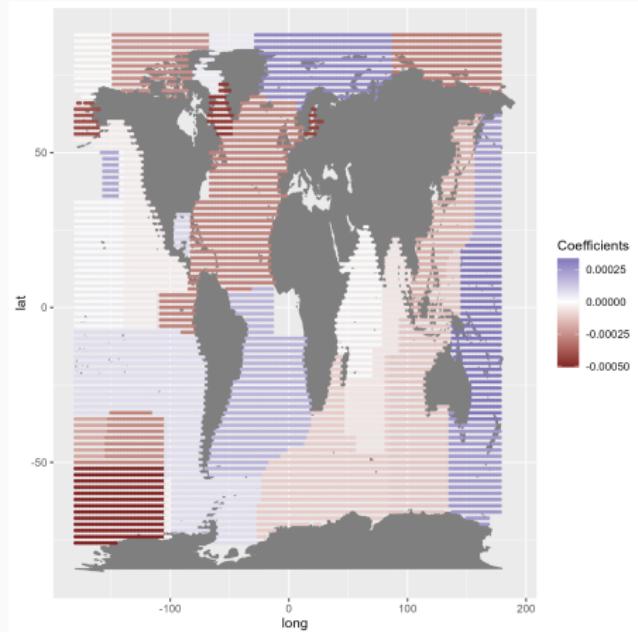


Figure 25: Coefficient plot of the final fused lasso model. Positive coefficient values are given in blue, negative values in red. Only values inside the 0.975 and 0.025 quantile range are shown.

Fused lasso - All model results

Table 2: Table of the MSE results for the different fused lasso models evaluated. The λ refers to value that was found in forward validation and used to fit the final model.

	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

Summary fused lasso results

- 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 fewer 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} .

Summary fused lasso results

- The coefficient plots reveal predictive connected areas with high negative values in the Baltic Sea and high positive values northeast of Canada.
- Since no sparsity is used, all areas obtain non-zero coefficient values, and some regions with high coefficient values are included in the lasso as well in the fused lasso.

7. Discussion & Conclusion

Discussion

- Our results suggest that precipitation can be predicted directly from SST.
- The overall predictability of precipitation in the CAB differed between the model selection and model evaluation phases.
- 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 validation approach was not optimal in its use of the data.
- We might have been too restrictive in exploiting the data or using data that became less relevant over time.

Discussion

- Possible other approaches:
- Allow for larger folds (introduces overlapping folds), or for the crossing of train and test in time (past is predicted with future values)
- Fit the full model with fewer data and discard data 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 actual situation well.
- Creating a weighted graph (will increase cost) 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 five almost completely spatially coherent clusters in the CAB.
- Standardizing the features yielded the best results for the lasso.
- On the hold-out data, the lasso fails to predict the peaks in precipitation but can predict low values well.

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.

Thank you

Thanks for your attention! :)

References i

- Arnold, Taylor B., and Ryan J. Tibshirani. 2020. *Genlasso: Path Algorithm for Generalized Lasso Problems.*
<https://CRAN.R-project.org/package=genlasso>.
- Arnold, Taylor B, and Ryan J Tibshirani. 2016. “Efficient Implementations of the Generalized Lasso Dual Path Algorithm.” *Journal of Computational and Graphical Statistics* 25 (1): 1–27.
- 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.

References ii

- Fahrmeir, Ludwig, Wolfgang Brachinger, Alfred Hamerle, and Gerhard Tutz. 1996. *Multivariate Statistische Verfahren*. Walter de Gruyter.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33 (1): 1–22. <https://doi.org/10.18637/jss.v033.i01>.
- 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.

References iii

- Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Vol. 2. Springer.
- 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.
- Lewis, Simon L, Paulo M Brando, Oliver L Phillips, Geertje MF Van Der Heijden, and Daniel Nepstad. 2011. “The 2010 Amazon Drought.” *Science* 331 (6017): 554–54.

- Schnaubelt, Matthias. 2019. "A Comparison of Machine Learning Model Validation Schemes for Non-Stationary Time Series Data." FAU Discussion Papers in Economics.
- Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–88.
- Tibshirani, Robert, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. 2005. "Sparsity and Smoothness via the Fused Lasso." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (1): 91–108.
- Tibshirani, Ryan J., and Jonathan Taylor. 2011. "The Solution Path of the Generalized Lasso." *The Annals of Statistics* 39 (3): 1335–71.

Van der Kooij, Anita J. 2007. *Prediction Accuracy and Stability of Regression with Optimal Scaling Transformations*. Leiden University.

Appendix

Clustering

k-means

- Our objective is to find k internally homogeneous and externally heterogeneous clusters (Hastie et al. 2009)
- Similarity is measured by the euclidean distance

$$d(x_i, x_{i'}) = \sum_{j=1}^p (x_{ij} - x_{i'j})^2 = \|x_i - x_{i'}\|^2 \quad (5)$$

k-means

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

$$W(C) = \frac{1}{2} \sum_{k=1}^K \sum_{C(i)=k} \sum_{C(i')=k} \|x_i - x_{i'}\|^2 = \sum_{k=1}^K N_k \sum_{C(i)=k} \|x_i - \bar{x}_k\|^2 \quad (6)$$

- 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
- If our data has k -clusters we expect W_k to be smaller than average W_k^* (tibshirani2001estimating):

$$Gap(k) = E\{\log(W_k^*)\} - \log(W_k). \quad (7)$$

- We choose k such that:

$$Gap(k) \geq Gap(k+1) - s_{k+1} \quad (8)$$

Gap statistic

- s_{k+1} is $sd_k \sqrt{1 + 1/B}$, and sd the standard deviation of $\log(W_k^*)$
- Before running the gap statistic and 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

PCA

- We use a principal component analysis to summarize the information in the p variables in a few new variables.
- The new variables (principal components) are linear combinations of the original ones and are orthogonal to each other.
- The first $k < p$ principal components explain the majority of the variance (Fahrmeir et al. 1996).

PCA

- We approximate our data with the rank-q linear model

$$f(\lambda) = \mu + V_q \lambda \quad (9)$$

- with μ a location vector in \mathbb{R}^p and V_q is a $p \times q$ matrix with columns being orthogonal unit vectors.
- λ is a q vector of parameters.
- We are trying to fit a hyperplane of rank q to the data.

PCA

- If we use least squares we have:

$$\min_{\mu, \{\lambda_i\}, V_q} \sum_{i=1}^N \|x_i - \mu - V_q \lambda_i\|^2 \quad (10)$$

- we have $\hat{\mu} = \bar{x}$ and $\hat{\lambda}_i = V_q^T(x_i - \bar{x})$, meaning we only need to find V_q .
- and we find V and by solving the singular value decomposition of X .

$$X = UDV^T \quad (11)$$

- Here UD gives us the principal components. And $H_q x_i = V_q V_q^T x_i$ is the orthogonal projection of a observation x_i to subspace that is spanned by the columns of V_q .

Scree Plot, PCA after centering

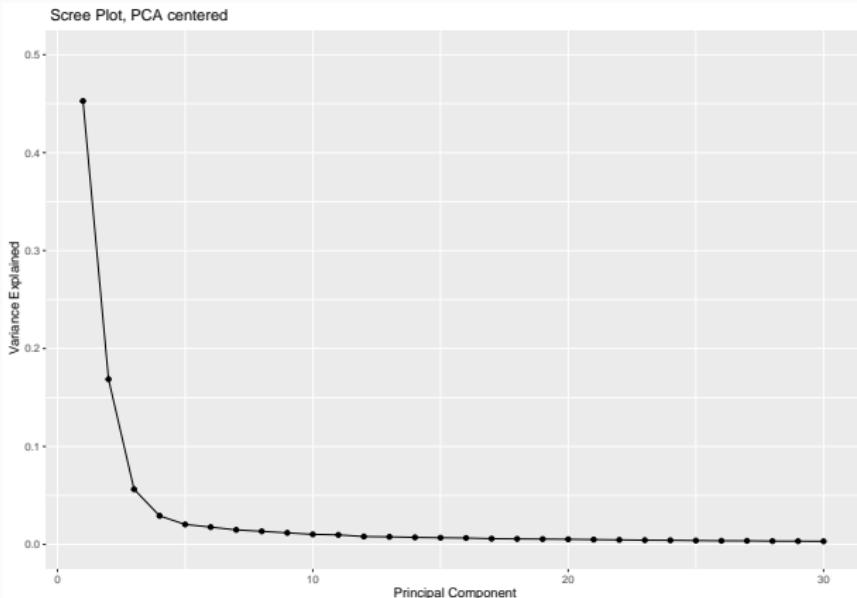


Figure 26: Scree plot of the Principal component analysis applied to the centered precipitation data.

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

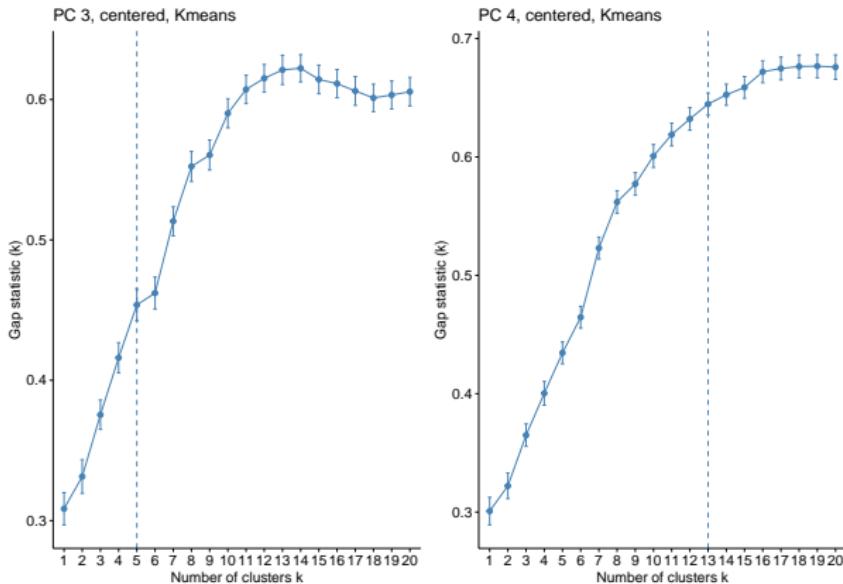


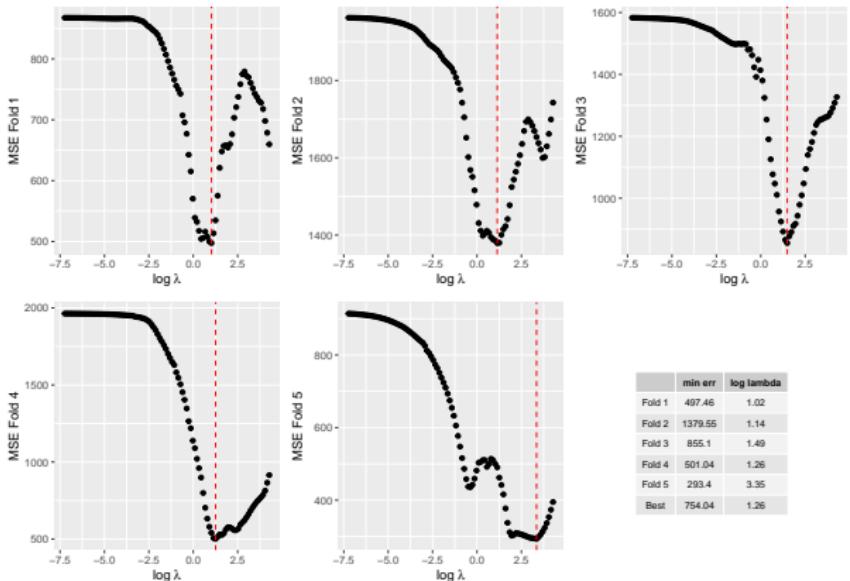
Figure 27: Results of the gap statistic when applying k-means on 3 (left) and 4 (right) principal components of the precipitation data.

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.

Lasso on original SST

MSE in each fold (Lasso on original SST)



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

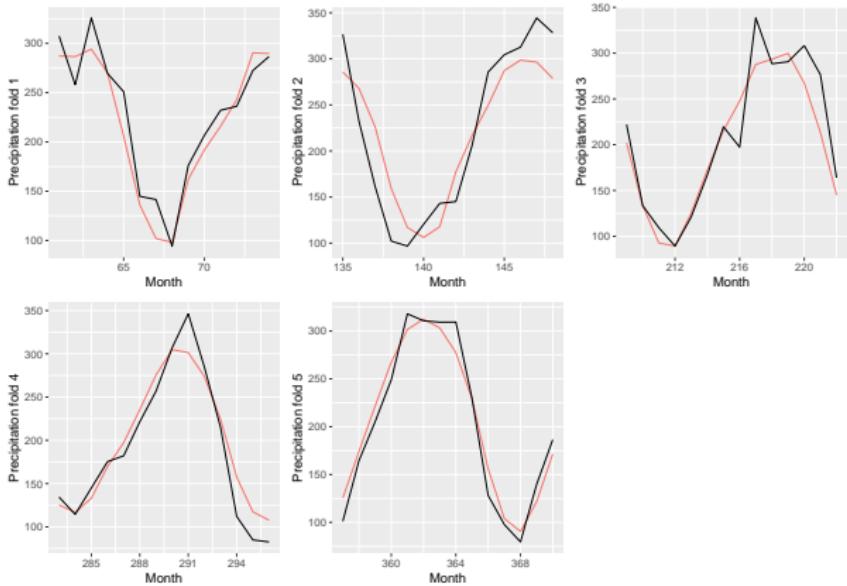


Figure 28: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set (Lasso on original SST)

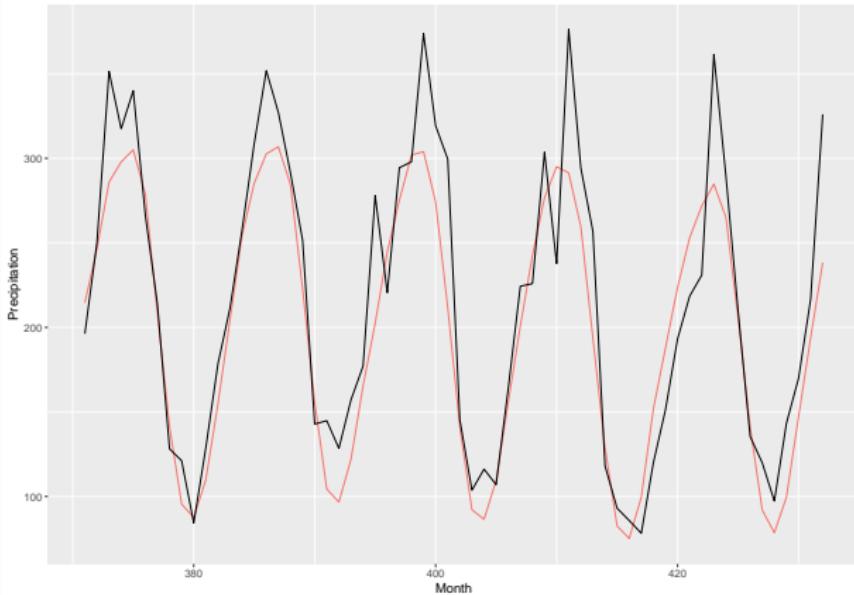


Figure 29: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE.

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

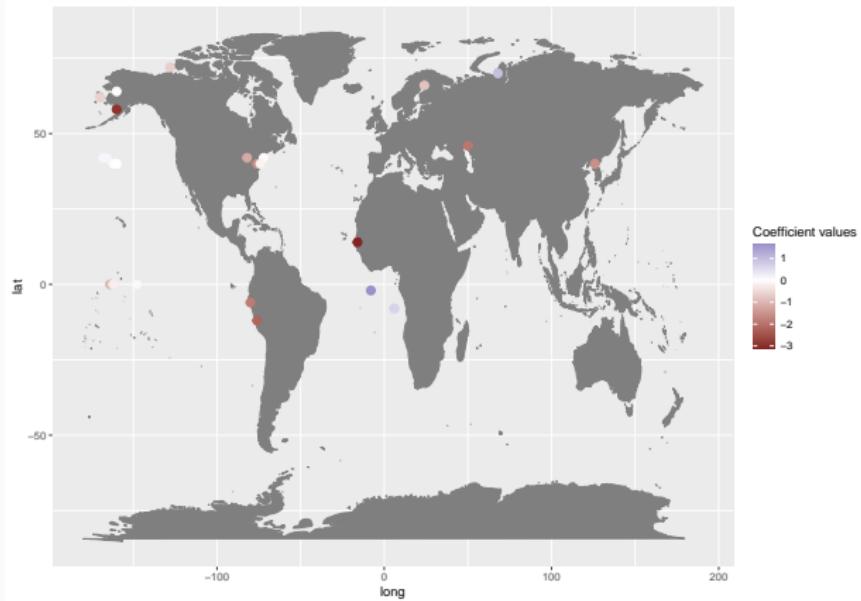
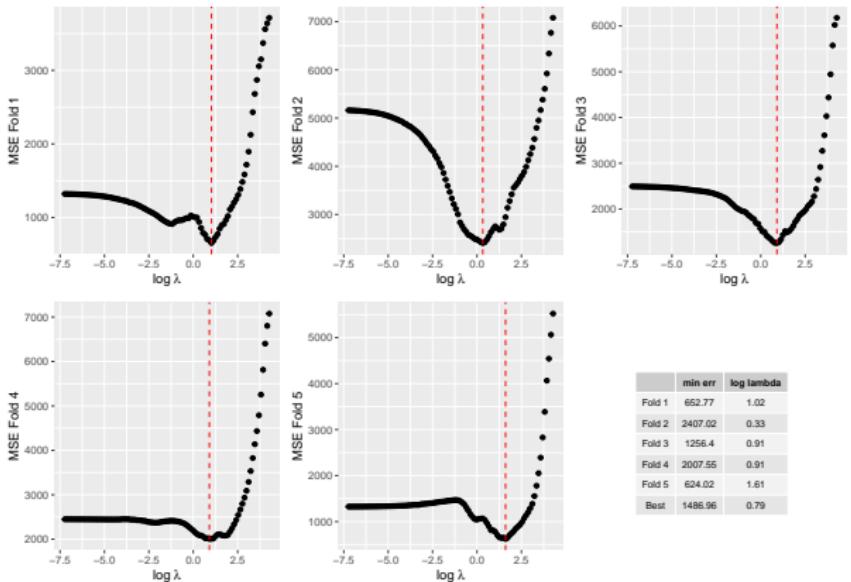


Figure 30: Coefficient plot of the full lasso model.

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)

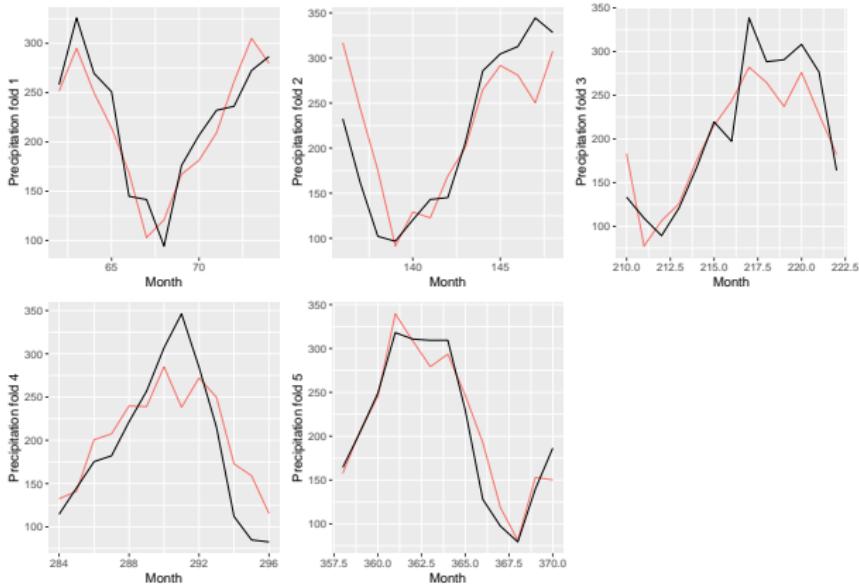


Figure 31: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set (Lasso differentiated)

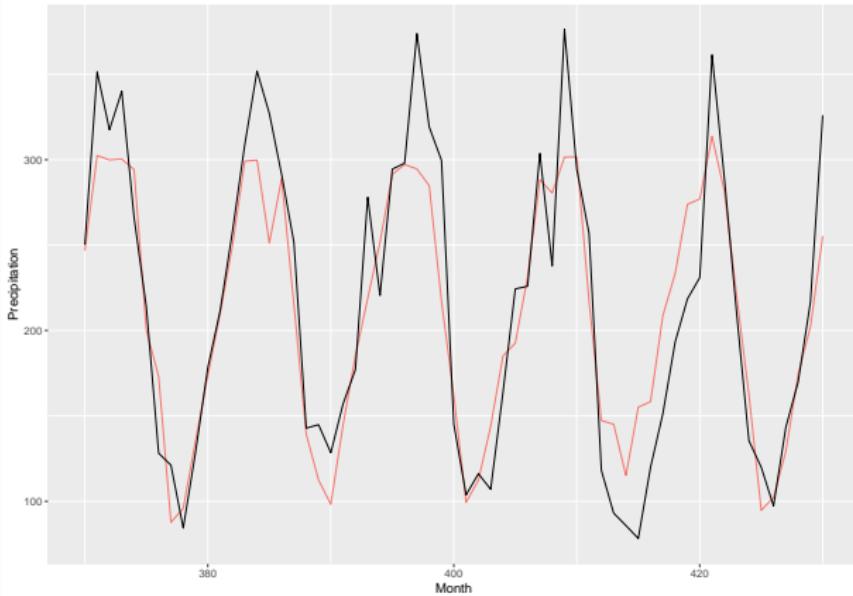


Figure 32: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

SST Regions chosen by the lasso (Lasso differentiated)

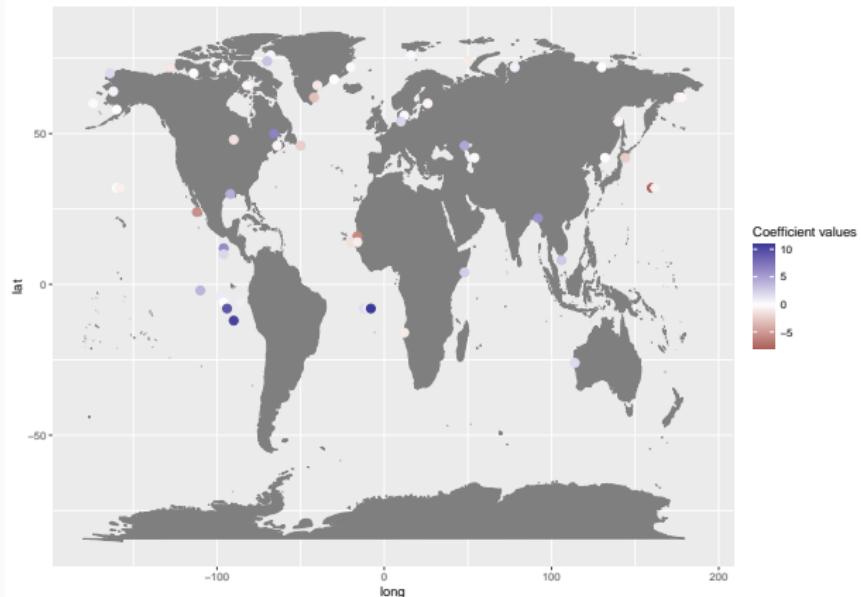
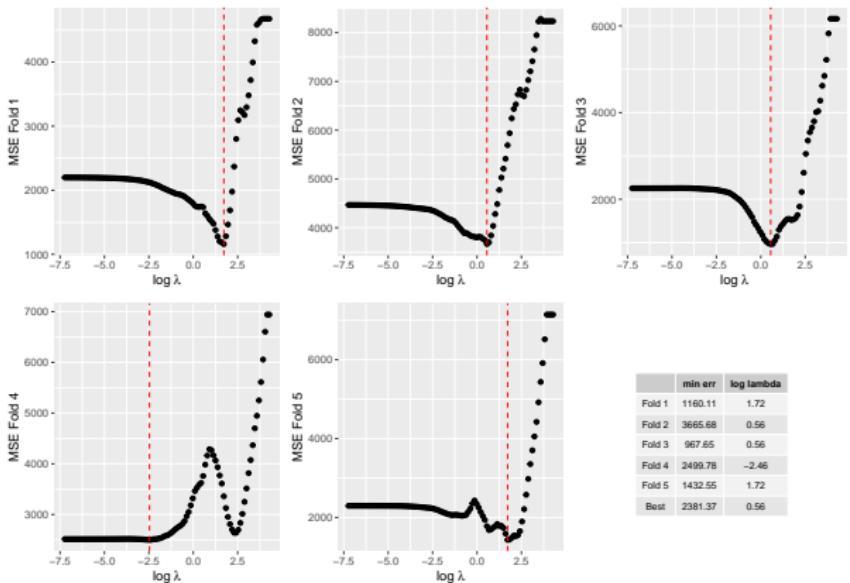


Figure 33: Coefficient plot of the full lasso model.

Lasso on de-seasonalized SST

MSE in each fold (Lasso de-seasonalized)



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

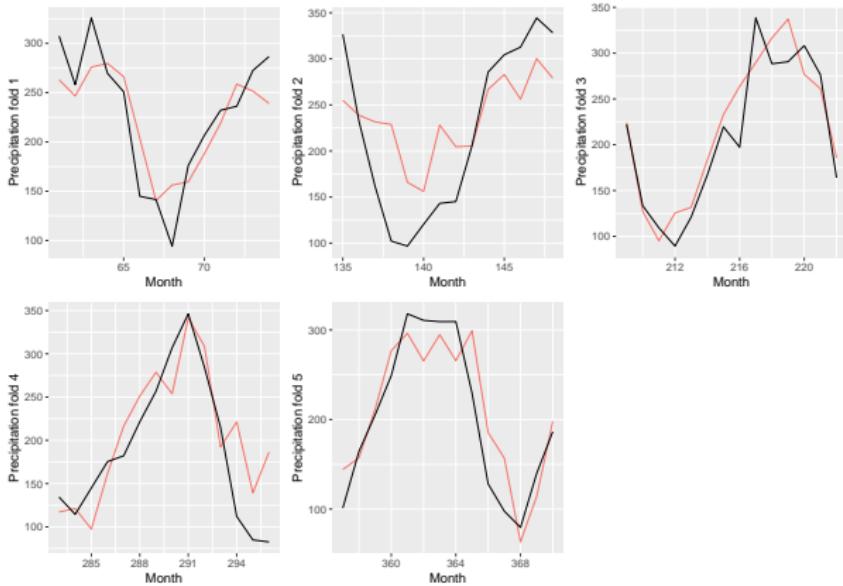


Figure 34: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

Predictions on External Test Set (Lasso de-seasonalized)

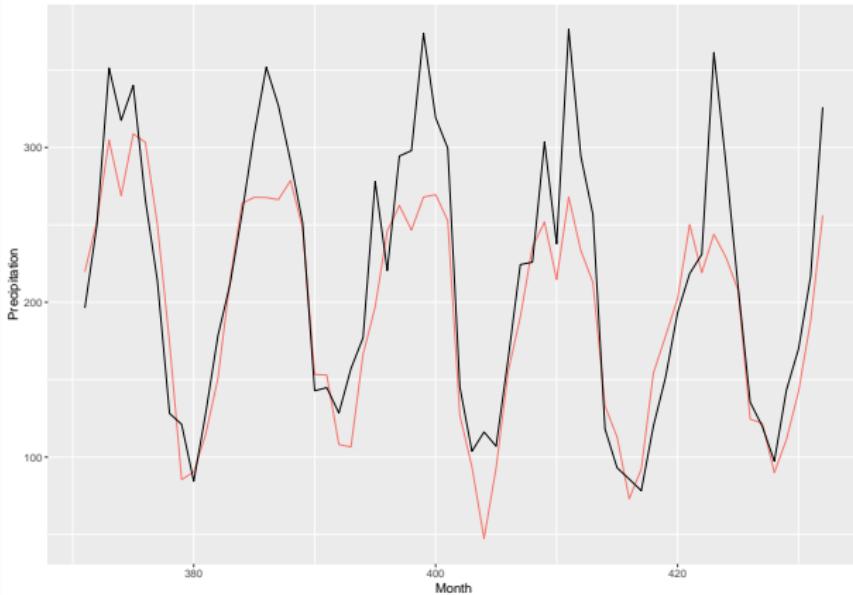


Figure 35: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

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

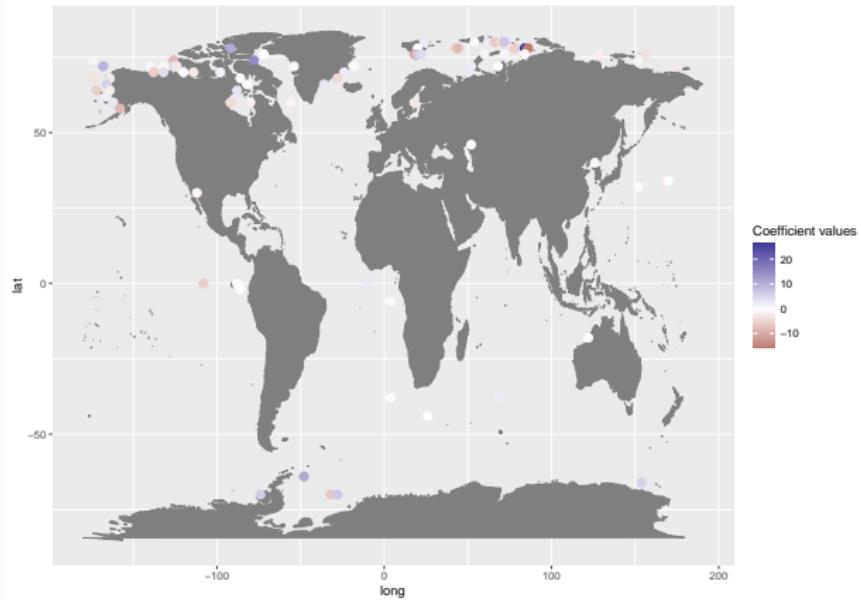


Figure 36: Coefficient plot of the full lasso model.

Fused lasso optimization

Fused lasso optimization

- 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 .
- We use the *genlasso* package and the *fusedlasso* function (Taylor B. Arnold and Tibshirani 2020)
- Let's consider the case when $X = I$ and $\text{rank}(X) = p$ (this is called the “signal approximator” case), the dual problem of (4) is then:

$$\hat{u} \in \arg \min_{u \in \mathbb{R}^m} \frac{1}{2} \|y - D^T u\|_2^2 \text{ subject to } \|u\|_\infty \leq \lambda. \quad (12)$$

Fused lasso optimization

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

$$\hat{\beta} = y - D^T \hat{u}. \quad (13)$$

- For general X and D with exploitable structure (as in our case), specialized implementations exist.
- The algorithm terminates when $\lambda = 0$ or a maximum number of steps is reached (we chose 5000, default is 20000).

Fused lasso optimization

- They use what they call algorithms 1 and 2.
- Algorithm 2 is a modification of 1 for general X , for sparse solves they add a small ridge penalty to X .
- Since D is the incidence matrix, they make use of DD^T is the Laplacian of the graph.
- The Laplacian linear systems are then solved using a Cholesky decomposition (see Section 4 and 5 in Taylor B. Arnold and Tibshirani (2016)).

Fused lasso with sub-graphs

Error lines (Fused lasso with sub-graphs)

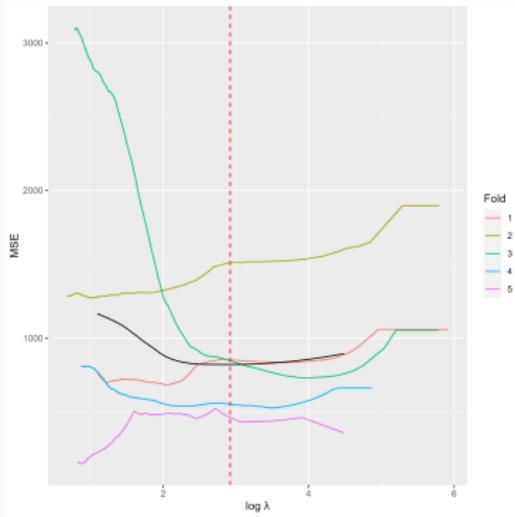


Figure 37: Error lines in the fused lasso forward validation. The model uses a graph including the sub-graphs. Each line represents one fold, the black line the mean for the common interval of the smoothed error lines. The red dashed-line shows the minimum of the mean.

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

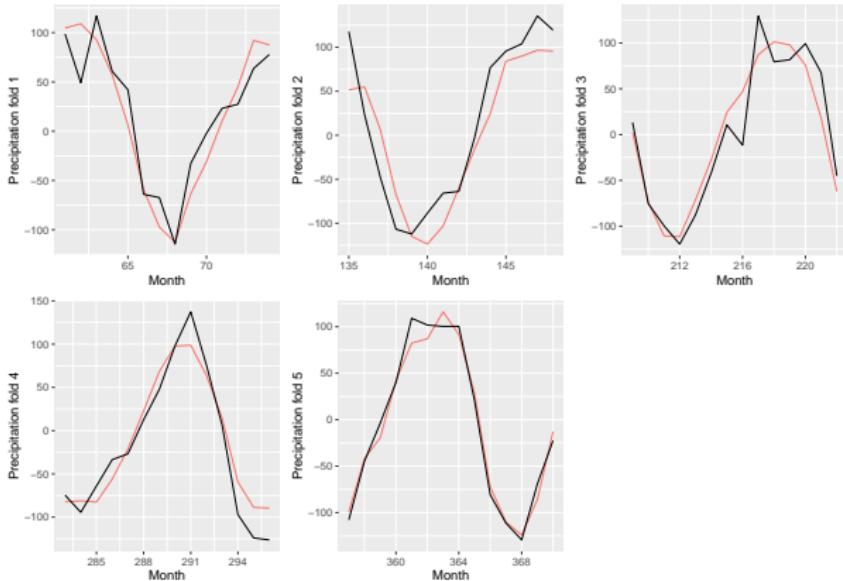


Figure 38: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

The predictions inside the folds are very similar to lasso without standardization (see 28), the same holds for the predictions from

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

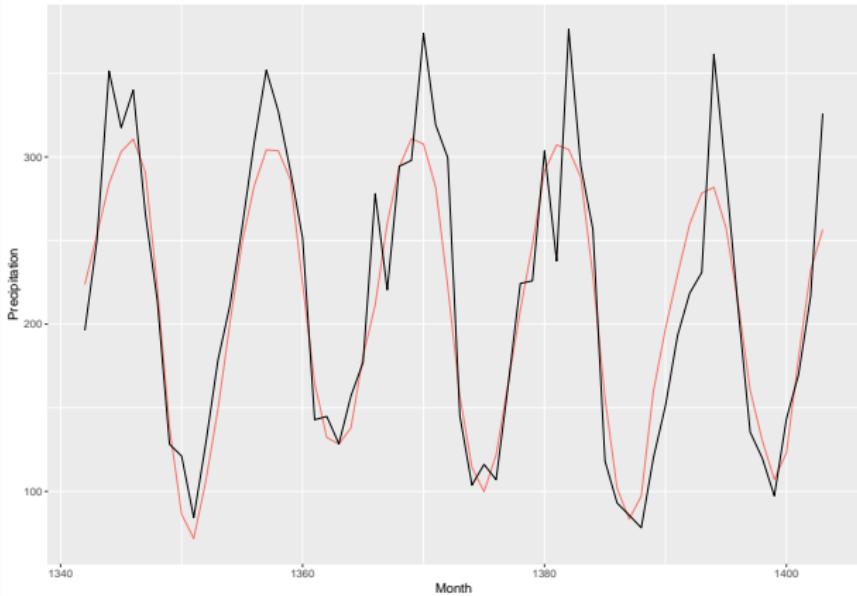


Figure 39: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

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

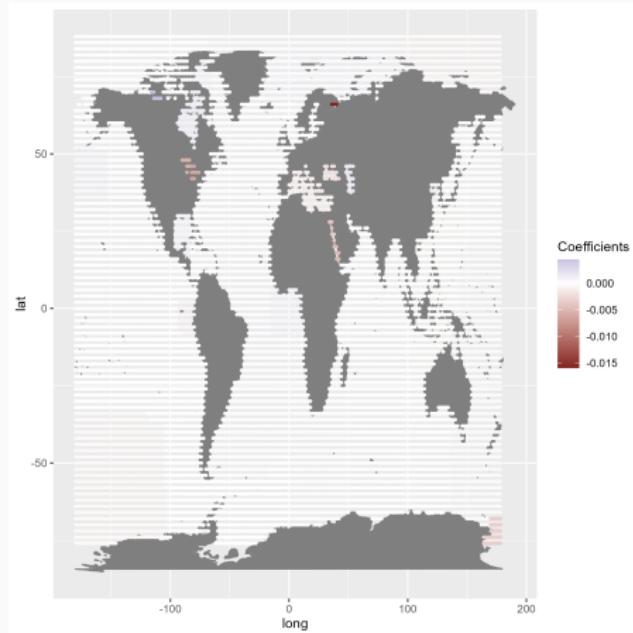


Figure 40: Coefficient plot of the final fused lasso model (fused lasso with sub-graphs). Positive coefficient values are given in blue, negative values in red.

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

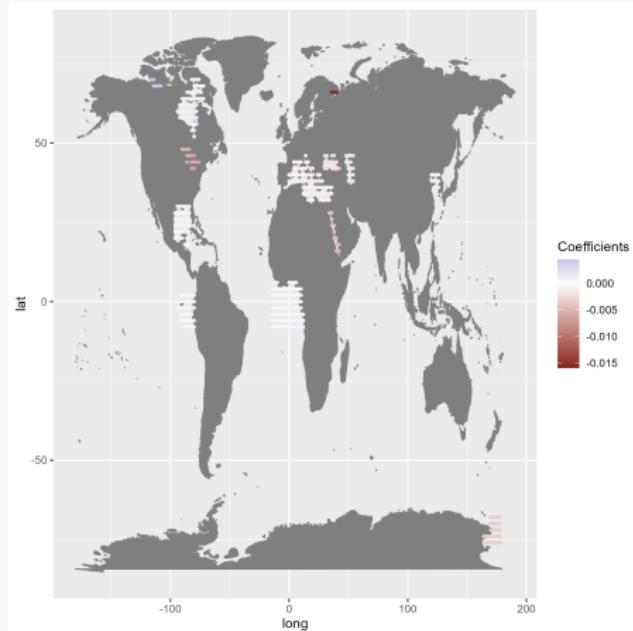


Figure 41: Coefficient plot of the final fused lasso model (fused lasso with sub-graphs). Positive coefficient values are given in blue, negative values in red. Only values outside the 97.5% and 2.5% quantile range are shown

**Fused lasso without sub-graphs,
gamma 0.1**

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

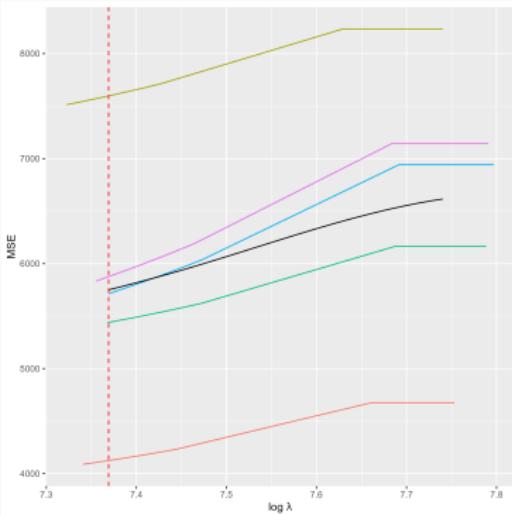


Figure 43: Error lines in the fused lasso forward validation. The model uses a graph without sub-graphs and $\gamma = 0.1$. Each line represents one fold, the black line the mean for the common interval of the smoothed error lines. The red dashed-line shows the minimum of the mean.

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

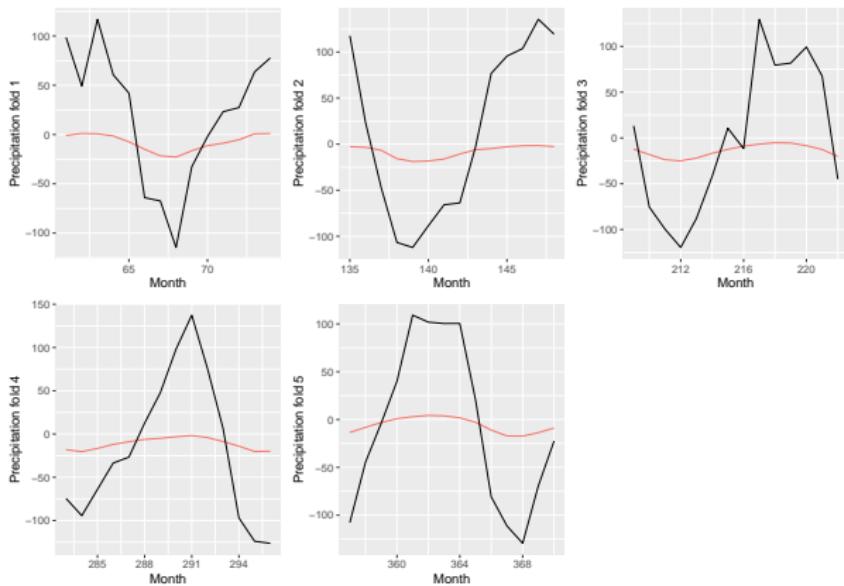


Figure 44: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

The predictions inside the folds are very similar to lasso without

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

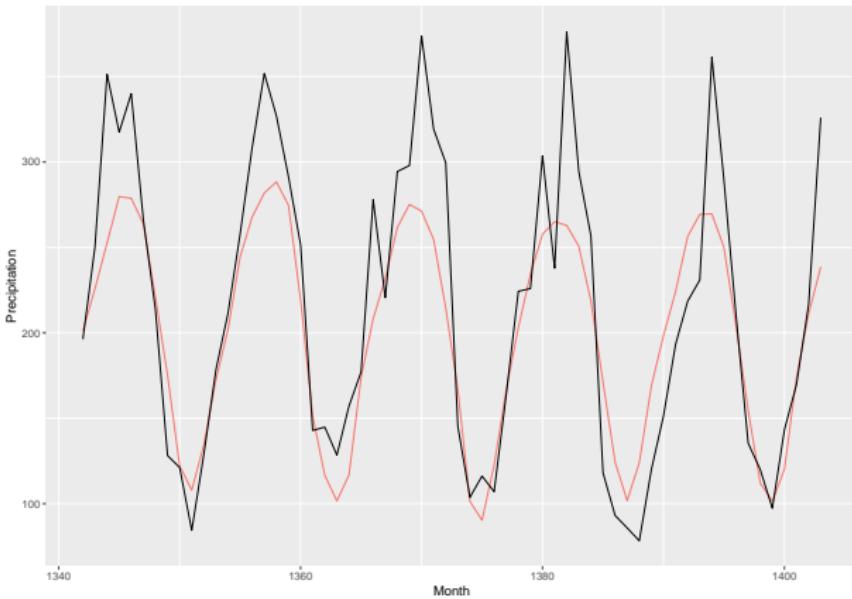


Figure 45: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

Coefficient plot of the final model (Fused lasso without sub-graphs, gamma 0.1)

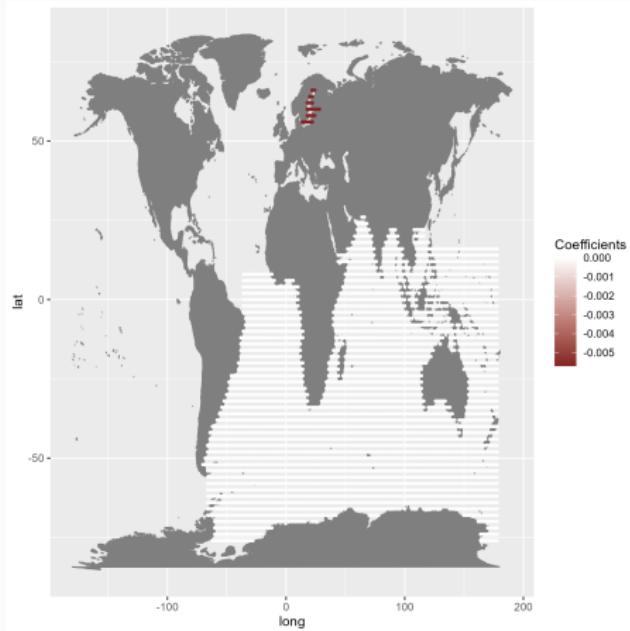


Figure 46: Coefficient plot of the final fused lasso model (fused lasso without sub-graphs and $\gamma = 0.1$). Positive coefficient values are given in blue, negative values in red.

**Fused lasso without sub-graphs,
gamma 0.05**

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

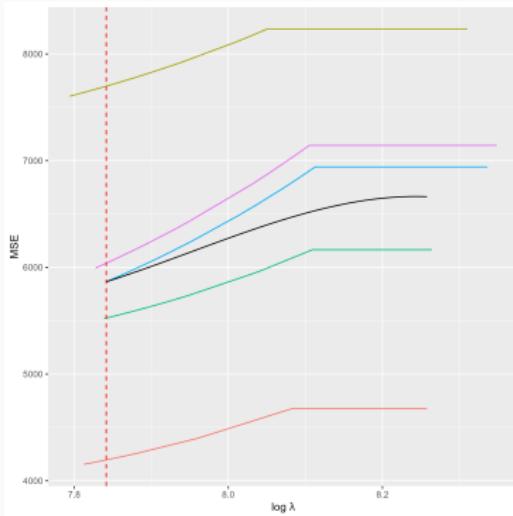


Figure 47: Error lines in the fused lasso forward validation. The model uses a graph without sub-graphs and $\gamma = 0.05$. Each line represents one fold, the black line the mean for the common interval of the smoothed error lines. The red dashed-line shows the minimum of the mean.

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

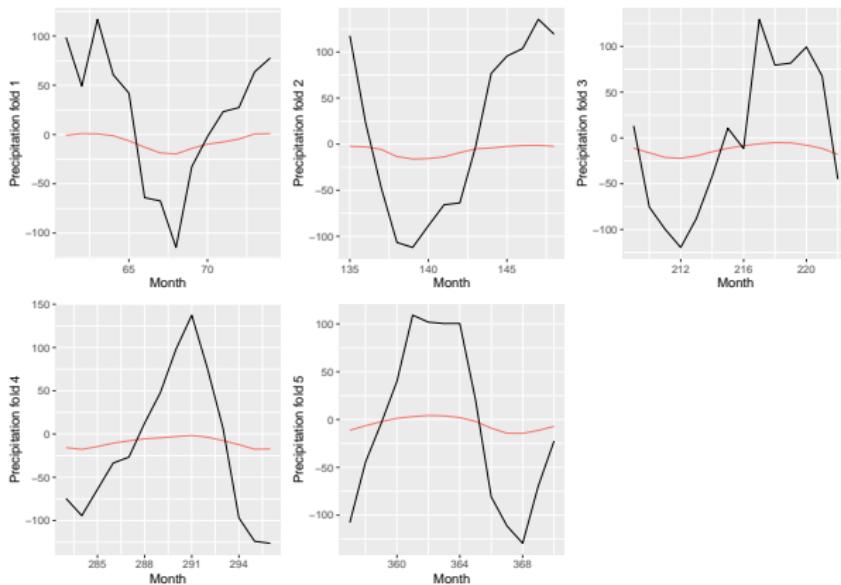


Figure 48: Precipitation prediction and target values in the test set in each fold. Predictions in red and target values in black.

The predictions inside the folds are very similar to lasso without

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

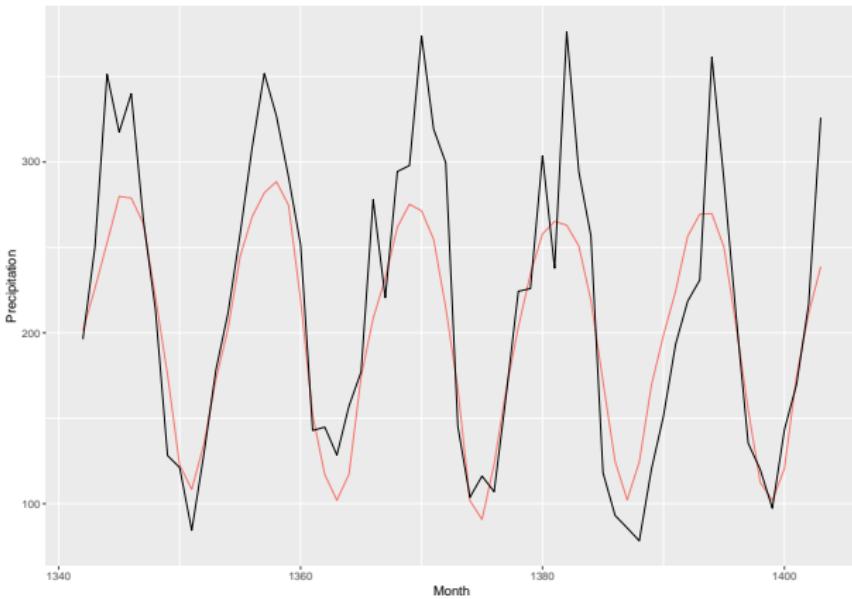


Figure 49: Precipitation prediction and target values in the validation set. Predictions in red and target values in black. The model was fitted on the full CV data with the lambda value that minimised the average MSE

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

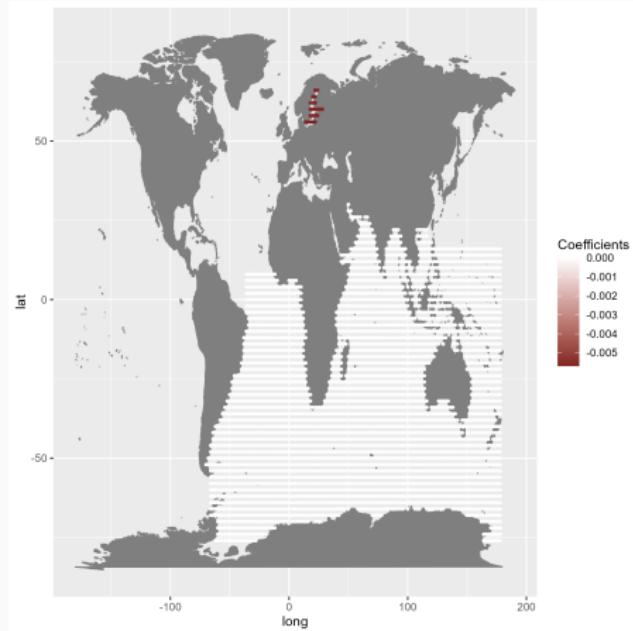


Figure 50: Coefficient plot of the final fused lasso model (fused lasso without sub-graphs and $\gamma = 0.1$). Positive coefficient values are given in blue, negative values in red.