Regularized censored regression with conditional heteroscedasticity

Jakob W. Messner,
Achim Zeileis

2017-12-17

Weather forecasts

Numerical Weather Prediction (NWP)

- Observations $\rightarrow$ estimate current atmospheric state.
- Simulate atmospheric processes with numerical models.

$\Rightarrow$ Compute future weather

Weather forecasts

Numerical Weather Prediction (NWP)

- Observations $\rightarrow$ estimate current atmospheric state.
- Simulate atmospheric processes with numerical models.

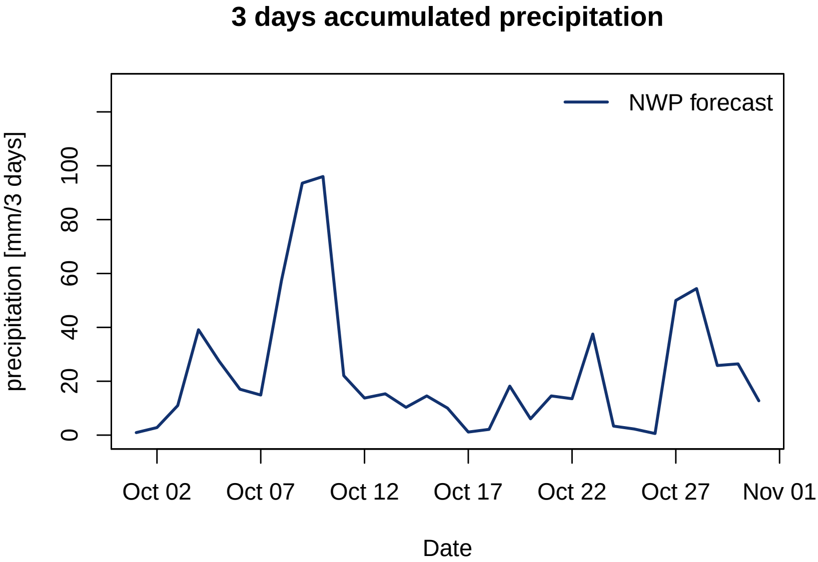
$\Rightarrow$ Compute future weather

Problems:

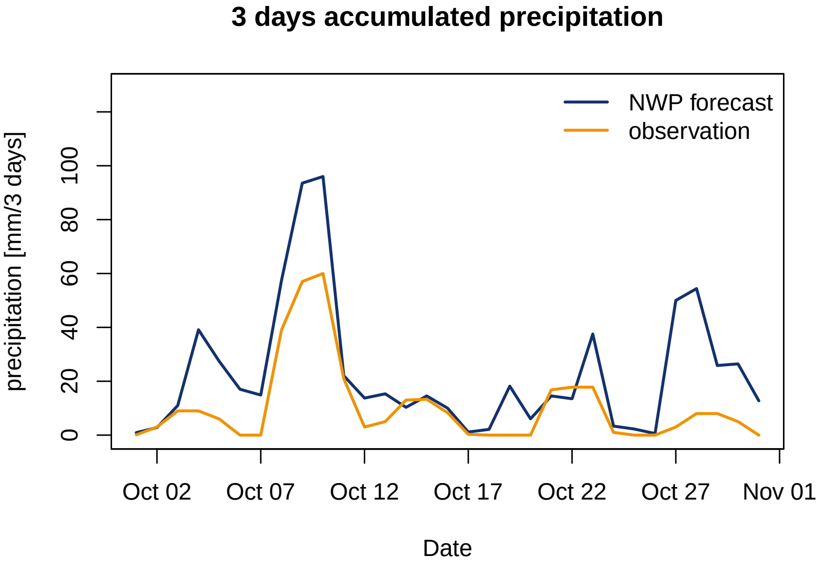
- Few observations
- Observation errors
- Not perfectly known atmospheric processes
- Unresolved processes

$\Rightarrow$ NWP errors

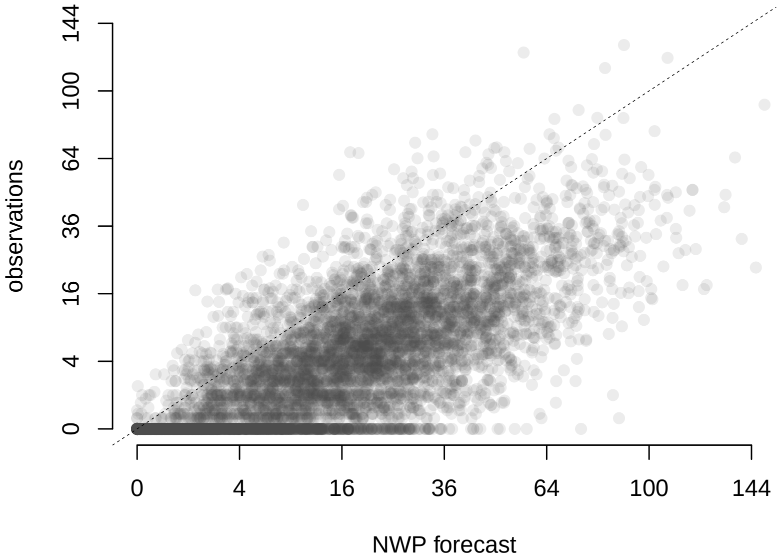
NWP errors


3 days accumulated precipitation
NWP forecast
precipitation [mm/3 days]
0
20
40
60
80
100
Oct 02
Oct 07
Oct 12
Oct 17
Oct 22
Oct 27
Nov 01
Date

NWP errors

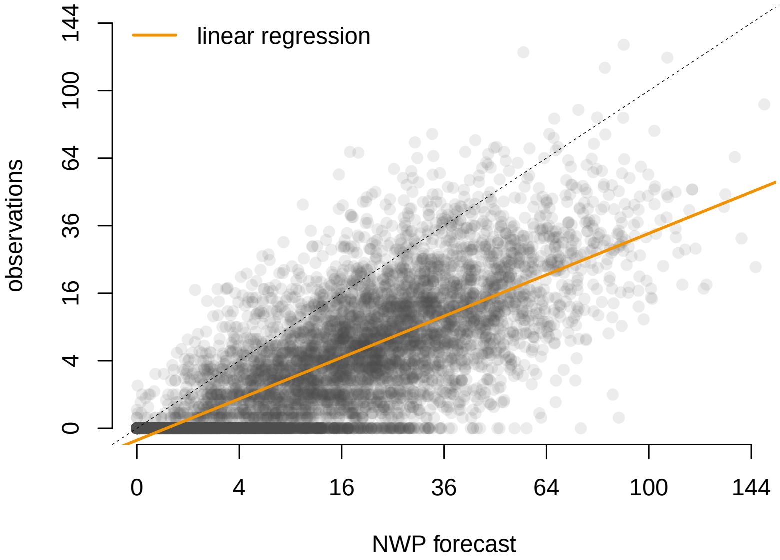

3 days accumulated precipitation
NWP forecast
observation
precipitation [mm/3 days]
0
20
40
60
80
100
Oct 02
Oct 07
Oct 12
Oct 17
Oct 22
Oct 27
Nov 01
Date

NWP errors



144
100
64
36
16
4
0
observations
0
4
16
36
64
100
144
NWP forecast

NWP errors



linear regression
observations
NWP forecast
0
4
16
36
64
100
144
0
4
16
36
64
100
144

Ensemble prediction

NWP error sources:

- Initial conditions
- Model formulations

Ensemble prediction

NWP error sources:

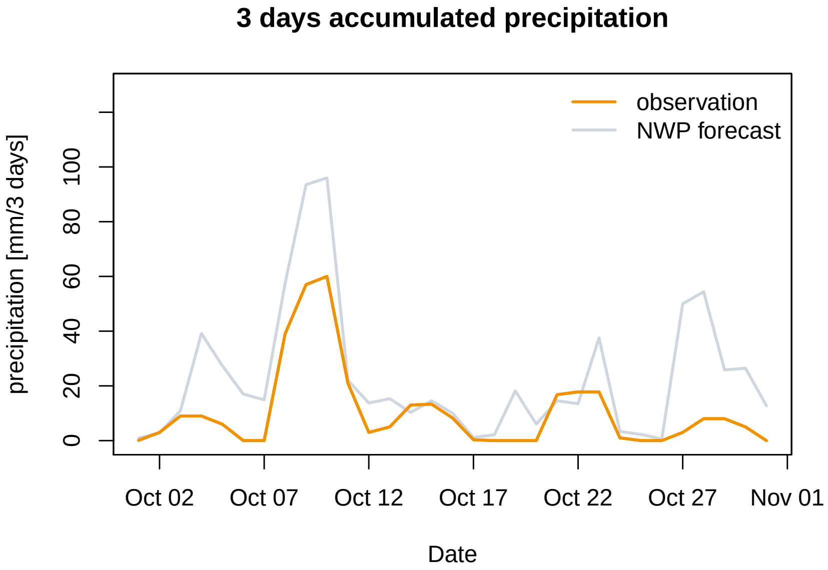
- Initial conditions
- Model formulations

Idea:

- Perturbed initial conditions
- Different model formulations

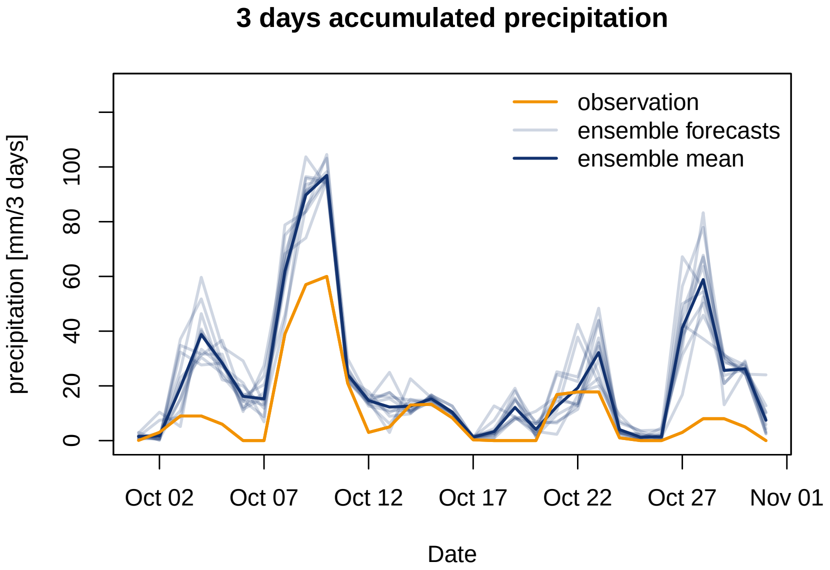
$\Rightarrow$ Compute different weather scenarios

NWP errors



3 days accumulated precipitation
observation
NWP forecast
precipitation [mm/3 days]
0
20
40
60
80
100
Oct 02
Oct 07
Oct 12
Oct 17
Oct 22
Oct 27
Nov 01
Date

NWP errors



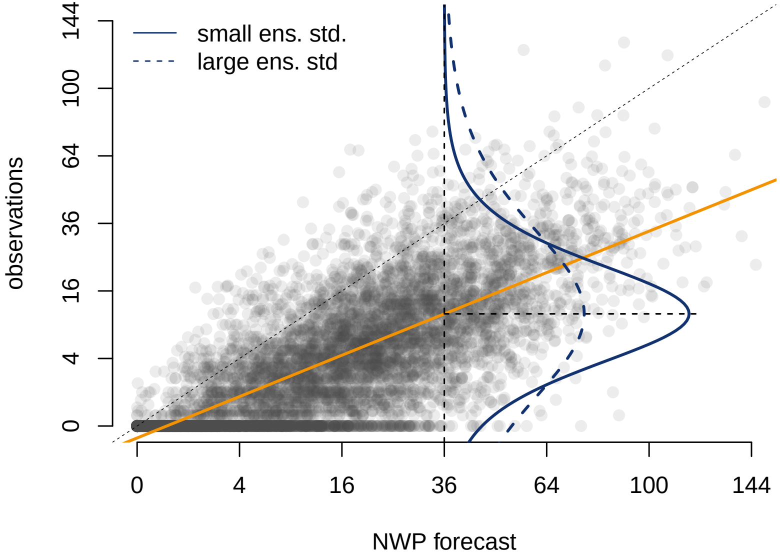
3 days accumulated precipitation
observation
ensemble forecasts
ensemble mean
precipitation [mm/3 days]
0
20
40
60
80
100
Oct 02
Oct 07
Oct 12
Oct 17
Oct 22
Oct 27
Nov 01
Date

Heteroscedastic censored regression

$$\begin{aligned} rain_i^* &\sim \mathcal{N}(\mu_i, \sigma_i^2) \\ \mu_i &= \beta_0 + \beta_1 * ensmean_i \\ \log(\sigma_i) &= \gamma_0 + \gamma_1 * \log(enssd_i) \end{aligned}$$

- $rain_i^*$: (latent) precipitation
- $ensmean_i$: ensemble mean forecast
- $enssd_i$: ensemble standard deviation
- $\beta_0, \beta_1, \gamma_0, \gamma_1$: regression coefficients

Heteroscedastic censored regression



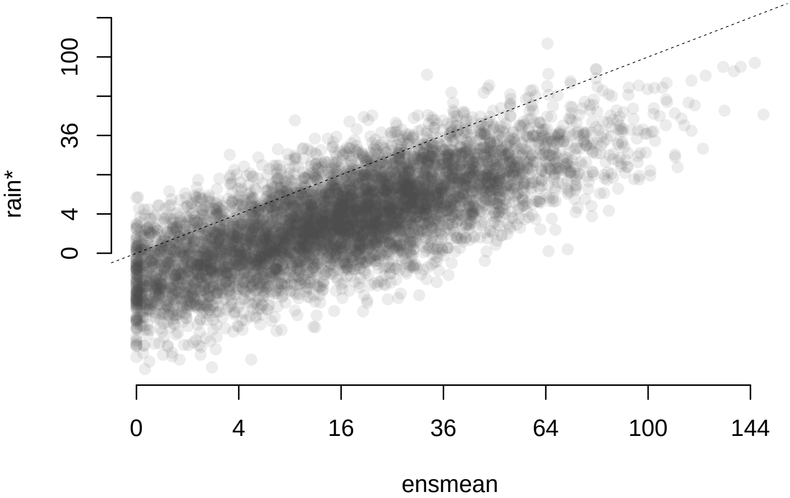
Heteroscedastic censored regression

$$\begin{aligned} rain_i^* &\sim \mathcal{N}(\mu_i, \sigma_i^2) \\ \mu_i &= \beta_0 + \beta_1 * ensmean_i \\ \log(\sigma_i) &= \gamma_0 + \gamma_1 * \log(enssd_i) \end{aligned}$$

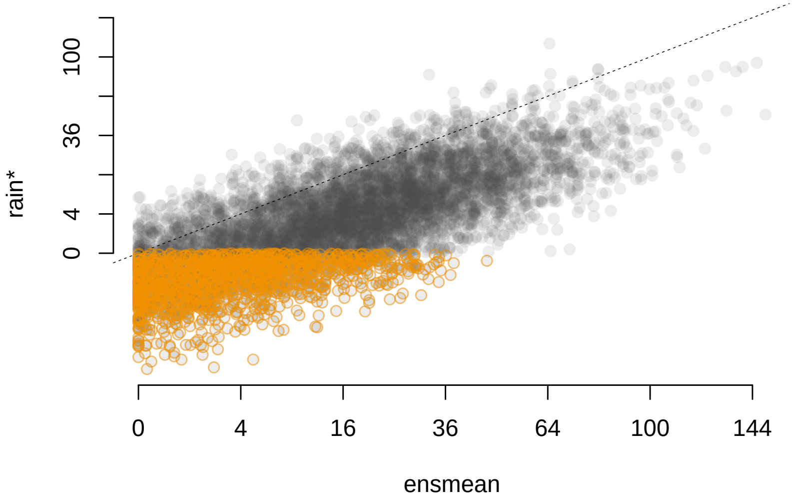
Consider non-negativity:

$$rain_i = \begin{cases} 0 & rain_i^* \leq 0 \\ rain_i^* & rain_i^* > 0 \end{cases}$$

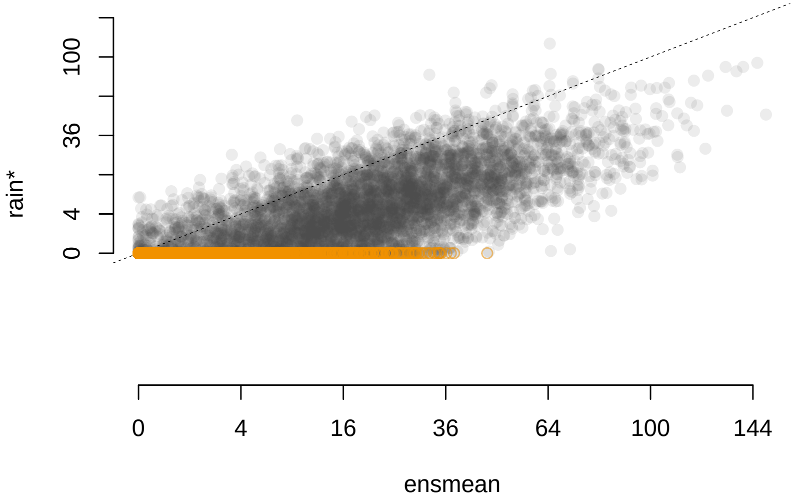
Heteroscedastic censored regression



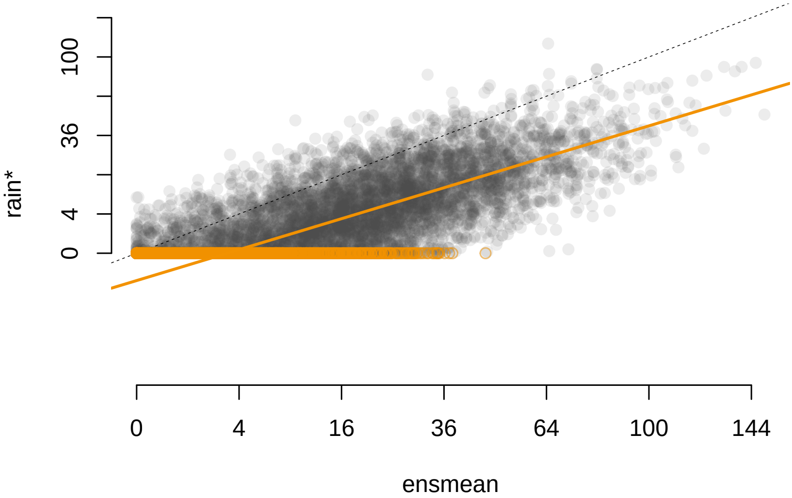
Heteroscedastic censored regression



Heteroscedastic censored regression



Heteroscedastic censored regression



Heteroscedastic censored regression

$$\begin{aligned} rain_i^* &\sim \mathcal{N}(\mu_i, \sigma_i^2) \\ \mu_i &= \beta_0 + \beta_1 * ensmean_i \\ \log(\sigma_i) &= \gamma_0 + \gamma_1 * \log(enssd_i) \end{aligned}$$

Consider non-negativity:

$$rain_i = \begin{cases} 0 & rain_i^* \leq 0 \\ rain_i^* & rain_i^* > 0 \end{cases}$$

Likelihood function:

$$L = \sum_{i=1}^{N} -\log(l_i) \qquad l_i = \begin{cases} \Phi\left(\frac{-\mu_i}{\sigma_i}\right) & rain_i = 0 \\ \phi\left(\frac{rain_i - \mu_i}{\sigma_i}\right) & rain_i > 0 \end{cases}$$

Input variables

So far: only ensemble mean and ensemble standard deviation of precipitation

Further potentially useful variables:

- ensemble predictions of other variables (e.g., temperature, wind)
- numerical predictions from other centers
- current observations
- transformations and interactions
- ...

Heteroscedastic censored regression

$$\begin{aligned} rain^* &\sim \mathcal{N}(\mu, \sigma^2) \\ \mu &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots = \mathbf{x}^\top \beta \\ \log(\sigma) &= \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \ldots = \mathbf{z}^\top \gamma \end{aligned}$$

- $rain^*$: (latent) precipitation
- $\mathbf{x}$: inputs for location
- $\mathbf{z}$: inputs for scale
- β, γ: regression coefficients

Heteroscedastic censored regression

$$\begin{aligned} rain^* &\sim \mathcal{N}(\mu, \sigma^2) \\ \mu &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots = \mathbf{x}^\top \beta \\ \log(\sigma) &= \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \ldots = \mathbf{z}^\top \gamma \end{aligned}$$

- $rain^*$: (latent) precipitation
- $\mathbf{x}$: inputs for location
- $\mathbf{z}$: inputs for scale
- β, γ: regression coefficients

Problem: How to select variables in $\mathbf{x}$ and $\mathbf{z}$?

LASSO penalization

$$\begin{aligned} rain^* &\sim \mathcal{N}(\mu, \sigma^2) \\ \mu &= \mathbf{x}^\top \beta \\ \log(\sigma) &= \mathbf{z}^\top \gamma \end{aligned}$$

Penalized negative log-likelihood:

$$L + \lambda \left(\sum_{j=1}^{P} |\beta_j| + \sum_{k=1}^{Q} |\gamma_k| \right)$$

- L: negative log-likelihood
- λ: penalization parameter
- P, Q: lengths of β and γ

LASSO penalization

$$\begin{aligned} rain^* &\sim \mathcal{N}(\mu, \sigma^2) \\ \mu &= \mathbf{x}^\top \beta \\ \log(\sigma) &= \mathbf{z}^\top \gamma \end{aligned}$$

Penalized negative log-likelihood:

$$L + \lambda \left(\sum_{j=1}^{P} |\beta_j| + \sum_{k=1}^{Q} |\gamma_k| \right)$$

Coordinate descent algorithm for regularization paths:
similar to `glmnet()` (Friedman et al., 2010)

- start with large λ: all coefficients zero
- decrease λ and optimize penalized log-likelihood
- efficient optimization by starting with previous solution

Gradient boosting

$$\begin{aligned} \textit{rain}^* &\sim \mathcal{N}(\mu, \sigma^2) \\ \mu &= \mathbf{x}^\top \beta \\ \log(\sigma) &= \mathbf{z}^\top \gamma \end{aligned}$$

1. standardize regressors (mean 0, variance 1)
2. initialize $\beta, \gamma = \mathbf{0}$ and $m = 0$
3. tentatively update β_j of regressor x_j that is most correlated to $-\partial L / \partial \mu$
4. tentatively update γ_j of regressor z_j that is most correlated to $-\partial L / \partial \sigma$
5. really update the coefficient that improves likelihood most
6. if $m < \mathit{mstop}$ update $m \leftarrow m + 1$ and repeat 3–5

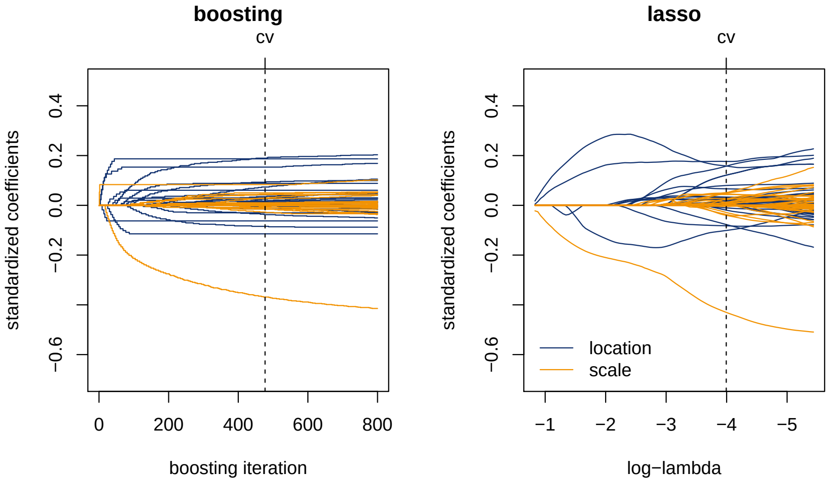
Gradient boosting (`gamboostLSS`)

$$
\begin{aligned}
rain^* &\sim \mathcal{N}(\mu, \sigma^2) \\
\mu &= \mathbf{x}^\top \beta \\
\log(\sigma) &= \mathbf{z}^\top \gamma
\end{aligned}
$$

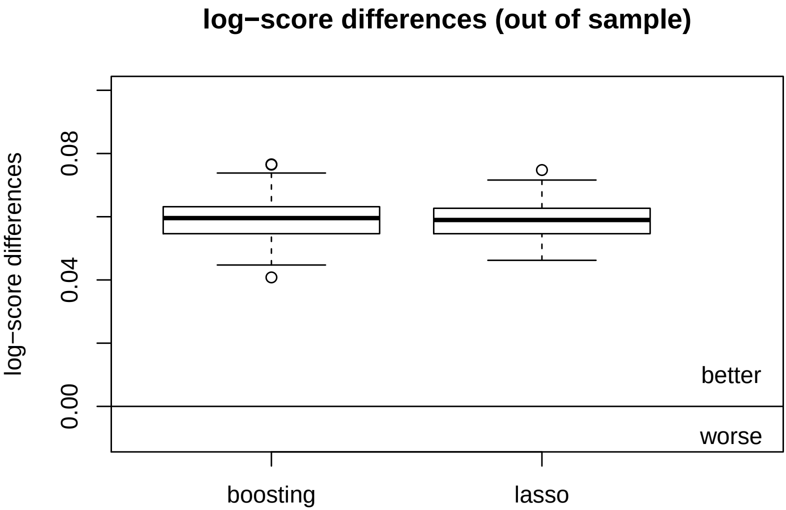
1. standardize regressors (mean 0, variance 1)
2. initialize $\beta, \gamma = \mathbf{0}$ and $m = 0$
3. if $m \leq mstop_\mu$:
   update β_j of regressor x_j that is most correlated to $-\partial L / \partial \mu$
4. if $m \leq mstop_\sigma$:
   update γ_k of regressor z_k that is most correlated to $-\partial L / \partial \sigma$
5. if $m < \max(mstop_\mu, mstop_\sigma)$ update $m \leftarrow m + 1$ and repeat 3–4

Problem: complex cross validation to optimize $mstop_\mu$ and $mstop_\sigma$

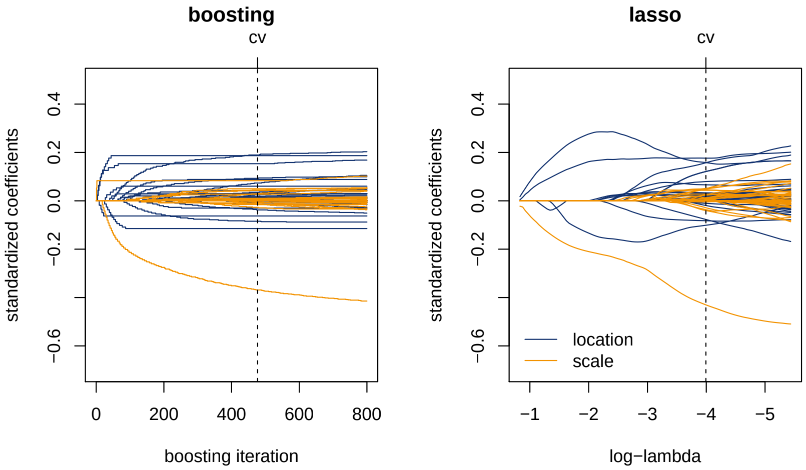
Boosting vs. LASSO



Boosting vs. LASSO



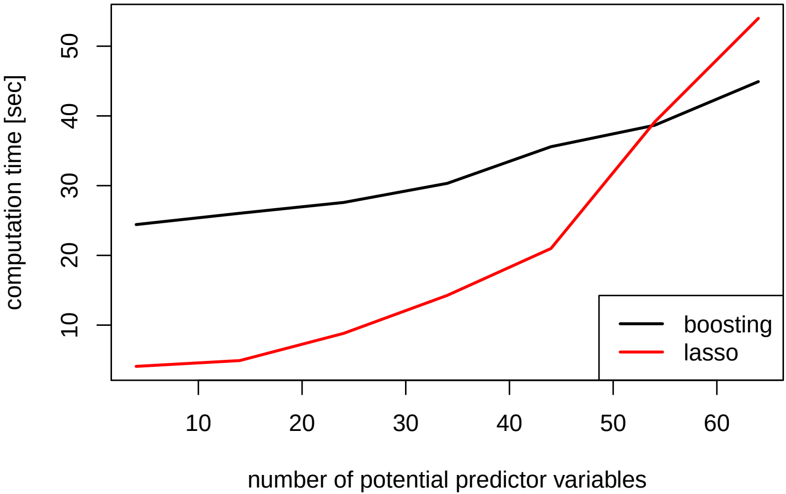
Boosting vs. LASSO



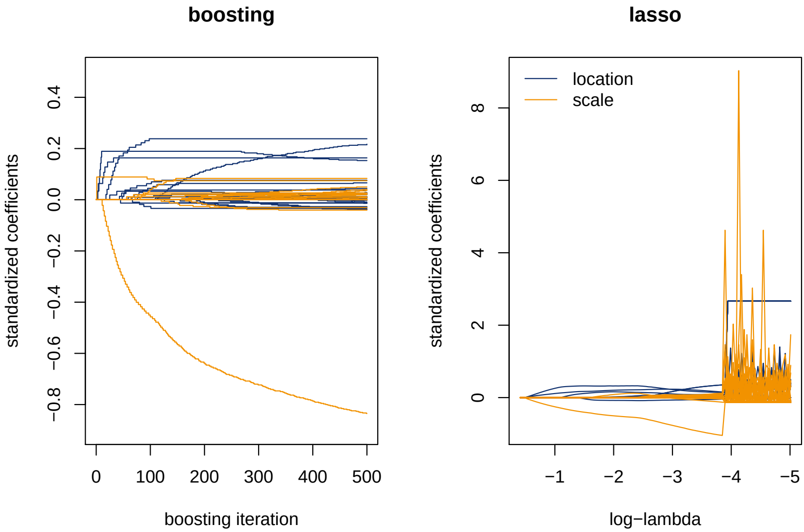
Approx. computation time (74 regressors, data length: 2755)

62 sec 70 sec

Boosting vs. LASSO



Boosting vs. LASSO



Summary

- both approaches select meteorologically reasonable variables and prevent overfitting
- LASSO
  - more efficient for smaller problems

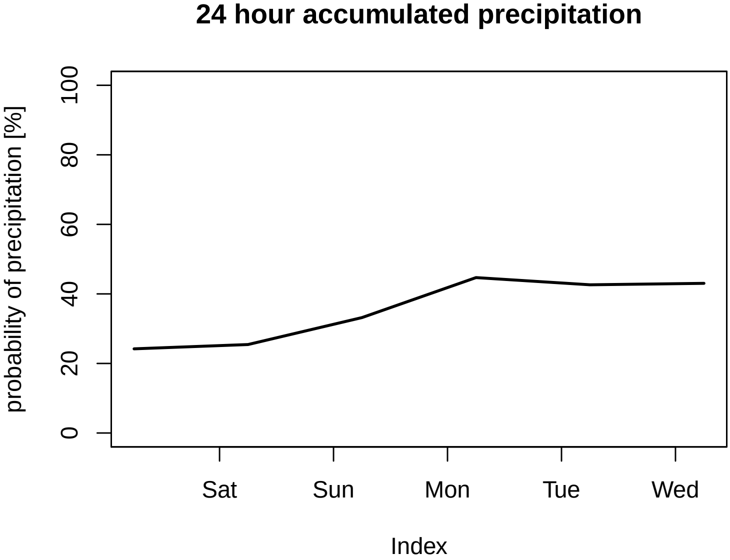
Summary

- both approaches select meteorologically reasonable variables and prevent overfitting
- LASSO
  - more efficient for smaller problems
- boosting
  - numerically more stable
  - more flexible: baselearners do not need to be linear models

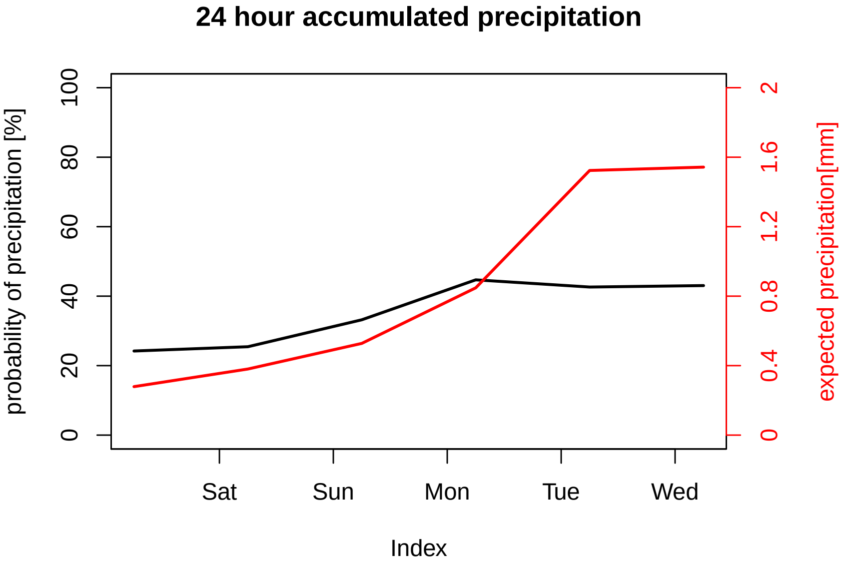
Summary

- both approaches select meteorologically reasonable variables and prevent overfitting
- LASSO
  - more efficient for smaller problems
- boosting
  - numerically more stable
  - more flexible: baselearners do not need to be linear models
- boosting implemented in CRAN package `crch`

London prediction



London prediction



London weather!

Friedman, J. H., T. Hastie, and R. Tibshirani, 2010: Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33 (1)**, 1–22.

Messner, J. W., G. J. Mayr, and A. Zeileis, 2016: Heteroscedastic censored and truncated regression with crch. *The R Journal*, **8 (1)**, 173–181, URL https://journal.r-project.org/archive/accepted/messner-mayr-zeileis.pdf.

Messner, J. W., G. J. Mayr, and A. Zeileis, 2017: Non-homogeneous boosting for predictor selection in ensemble post-processing. *Monthly Weather Review*, **145 (1)**, 137–147, doi:10.1175/MWR-D-16-0088.1.

jwmm@dtu.dk