



# Automatisierte Probabilistische (Süd-)föhnvorhersage

mittels statistischem Postprocessing

Reto Stauffer und Georg J. Mayr

[reto.stauffer@uibk.ac.at](mailto:reto.stauffer@uibk.ac.at)

# Einleitung

## **Bisherige Ansätze:**

- Widmer (1966), Courvoisier und Gutermann (1971),  
lineare Regression

# Einleitung

## **Bisherige Ansätze:**

- Widmer (1966), Courvoisier und Gutermann (1971), lineare Regression
- Sprenger et al. (2017), Adaptive Boosting mit COSMO Analyse, Nowcasting

# Einleitung

## **Bisherige Ansätze:**

- Widmer (1966), Courvoisier und Gutermann (1971), lineare Regression
- Sprenger et al. (2017), Adaptive Boosting mit COSMO Analyse, Nowcasting
- Zweifel (2016), Postprocessing mit logistischer Regression, Vorhersage +15/+39 Altdorf/Piotta

# Einleitung

## **Bisherige Ansätze:**

- Widmer (1966), Courvoisier und Gutermann (1971), lineare Regression
- Sprenger et al. (2017), Adaptive Boosting mit COSMO Analyse, Nowcasting
- Zweifel (2016), Postprocessing mit logistischer Regression, Vorhersage +15/+39 Altdorf/Piotta

## **Unsere Studie:**

- Angelehnt an Zweifel (2016)
- Allgemeinere Formulierung
- Weitere Standorte
- Längerer Vorhersagehorizont

# Einleitung

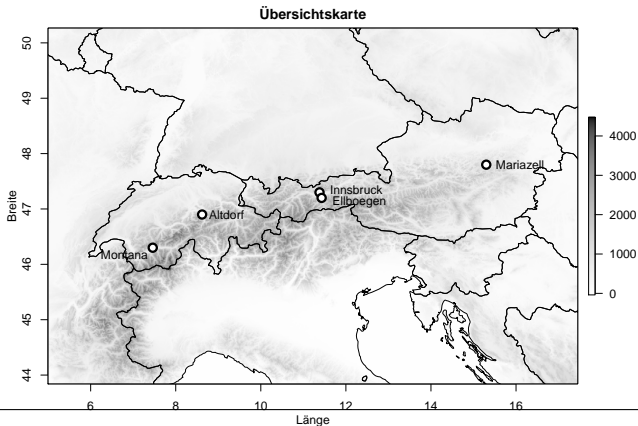
## **Ziel der Studie:**

- Vollautomatische probabilistische (Süd-)föhnvorhersage
- 3-stündig bis 6 Tage in die Zukunft
- Einfach adaptierbar auf neue Standorte

# Einleitung

## Ziel der Studie:

- Vollautomatische probabilistische (Süd-)föhnvorhersage
- 3-stündig bis 6 Tage in die Zukunft
- Einfach adaptierbar auf neue Standorte



# Datengrundlage

## **Beobachtungsdaten:**

- Südföhnklassifikation mittels Gaussian Mixture-Model
- Aggregation auf 3 h Intervalle ( $\pm 1.5$  h)
- Event: mindestens 1/3 Südföhn diagnostiziert

# Datengrundlage

## Beobachtungsdaten:

- Südföhnklassifikation mittels Gaussian Mixture-Model
- Aggregation auf 3 h Intervalle ( $\pm 1.5$  h)
- Event: mindestens 1/3 Südföhn diagnostiziert

Standort	$\bar{\pi}_{3h}$
Montana	14.8 %
Altdorf	7.3 %
Ellbögen	20.1 %
Innsbruck UNI	4.5 %
Mariazell	16.0 %

# Datengrundlage

## **Modellvorhersagen:**

- ECMWF HRES, 0000 UTC Lauf bis +6 d
- 6 Jahre (2012, ..., 2017)

# Datengrundlage

## **Modellvorhersagen:**

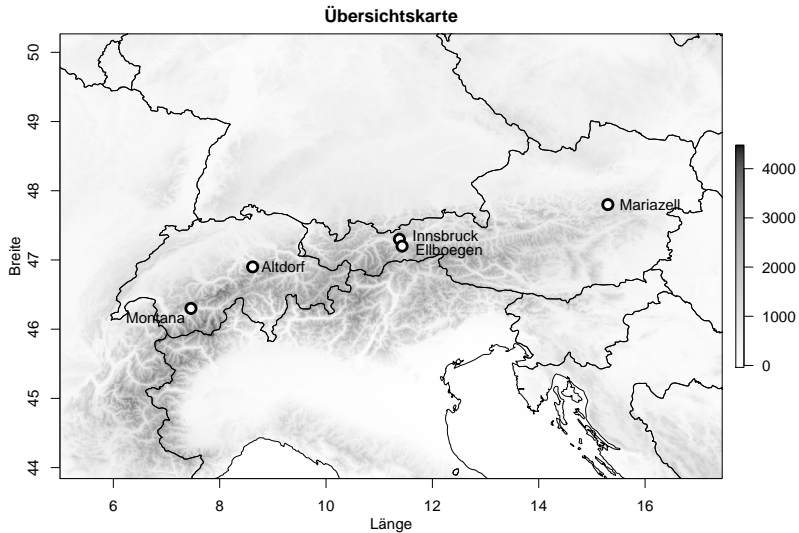
- ECMWF HRES, 0000 UTC Lauf bis +6 d
- 6 Jahre (2012, . . . , 2017)
- Direkte Grössen am Stationsstandort
- Abgeleitete Grössen:  
Stabilität, Schichtdicken, Scherung, . . .

# Datengrundlage

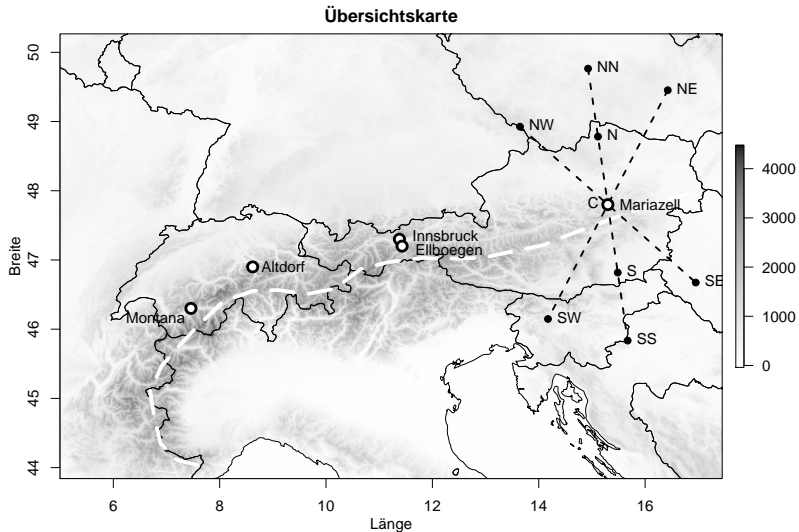
## **Modellvorhersagen:**

- ECMWF HRES, 0000 UTC Lauf bis +6 d
- 6 Jahre (2012, . . . , 2017)
- Direkte Grössen am Stationsstandort
- Abgeleitete Grössen:  
Stabilität, Schichtdicken, Scherung, . . .
- Zeitliche Gradienten (-3 h/0 h/+3 h)
- Räumliche Gradienten sowie  
Mittelwerte/Minima/Maxima Süd/Nord, . . .

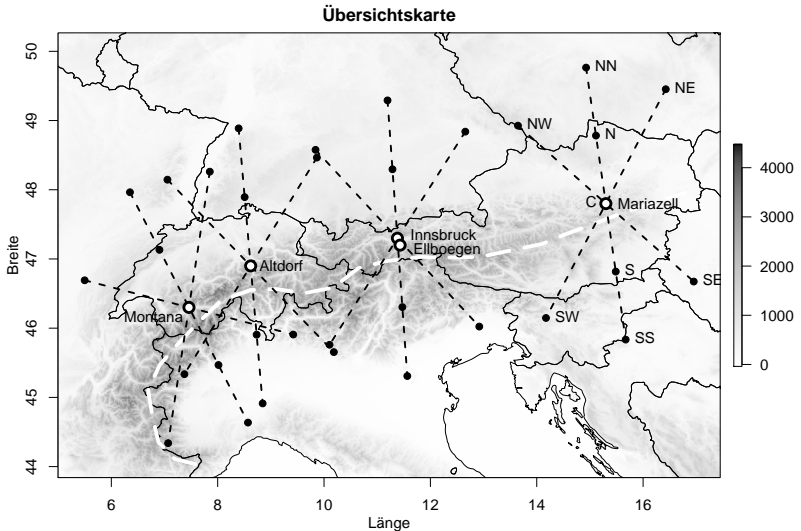
# Datengrundlage



# Datengrundlage



# Datengrundlage



⇒ Insgesamt >400 Kovariablen.

# Methodik

## Logistische Regression:

$$\text{logit}(\pi) = \beta_0 + \beta_1 X_1 + \dots + \beta_P X_P \quad (1)$$

# Methodik

## Logistische Regression:

$$\text{logit}(\pi) = \beta_0 + \beta_1 X_1 + \dots + \beta_P X_P \quad (1)$$

## Automatische Variablenselektion

- Lasso
- Gradient Boosting

# Methodik

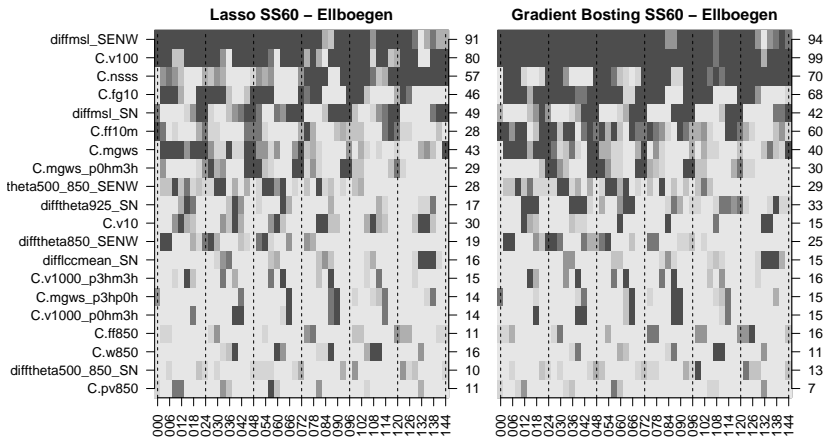
## Logistische Regression:

$$\text{logit}(\pi) = \beta_0 + \beta_1 X_1 + \dots + \beta_P X_P \quad (1)$$

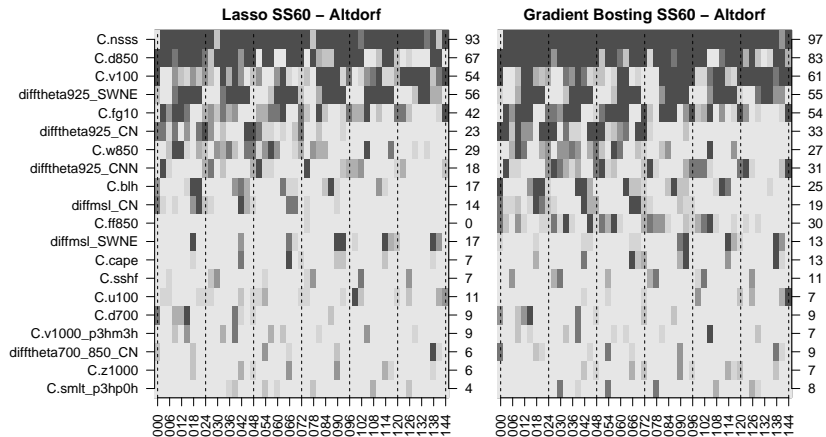
## Automatische Variablenselektion

- Lasso
- Gradient Boosting
- Stability Selection

# Variablenselektion



# Variablenselektion



# Resultate

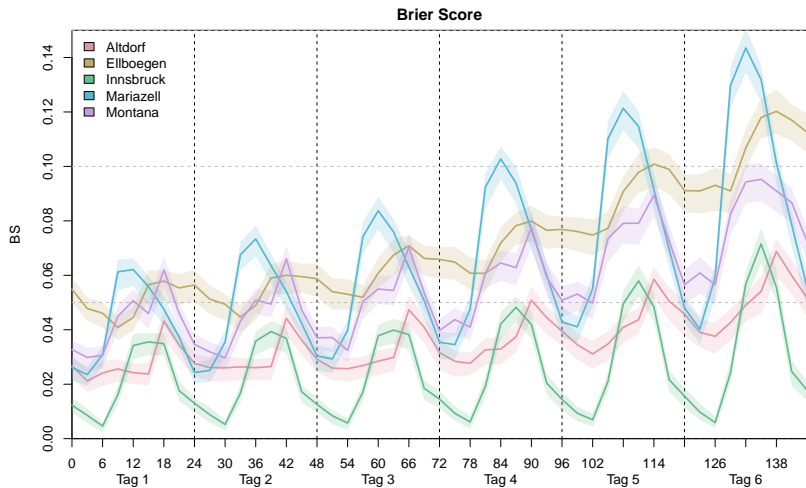
## Brier Score

$$BS = \frac{1}{N} \sum_{i=1}^N (\pi_i - o_i)^2 \quad (2)$$

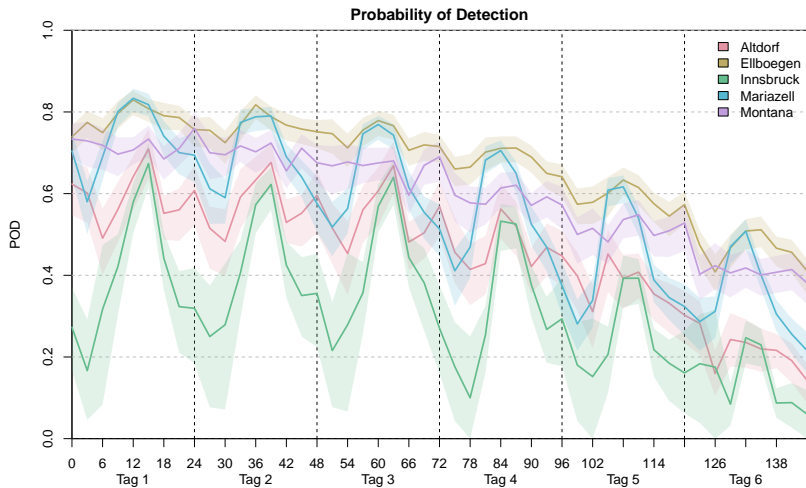
## Probability of Detection

$$POD = P(\pi_i \geq 0.5 \mid y_i = 1) \quad (3)$$

# Resultate

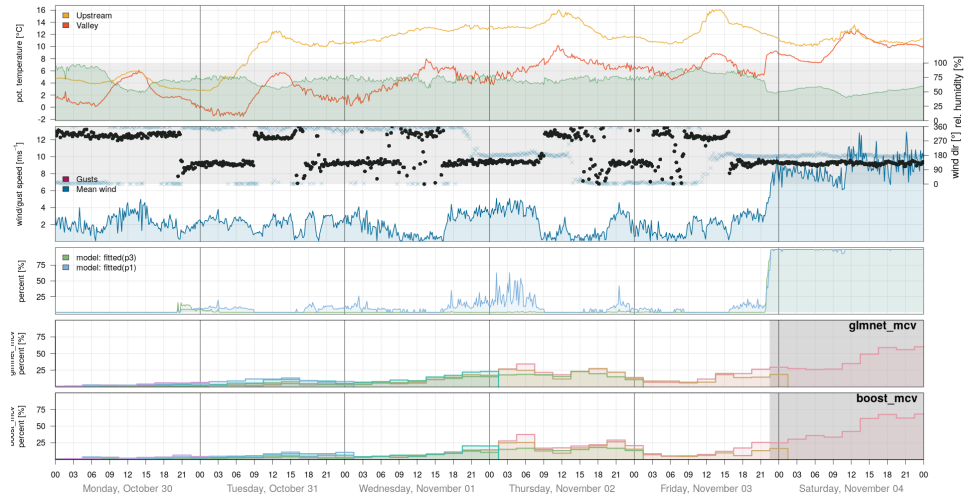


# Resultate



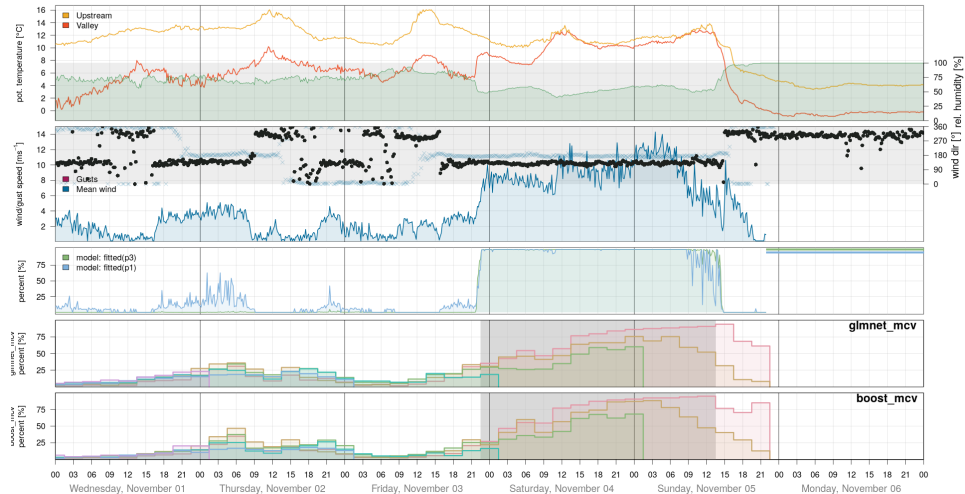
# Fallstudie: Ellbögen

Foehn Forecast Plot, Ellbogen  
2017-10-30 - 2017-11-05



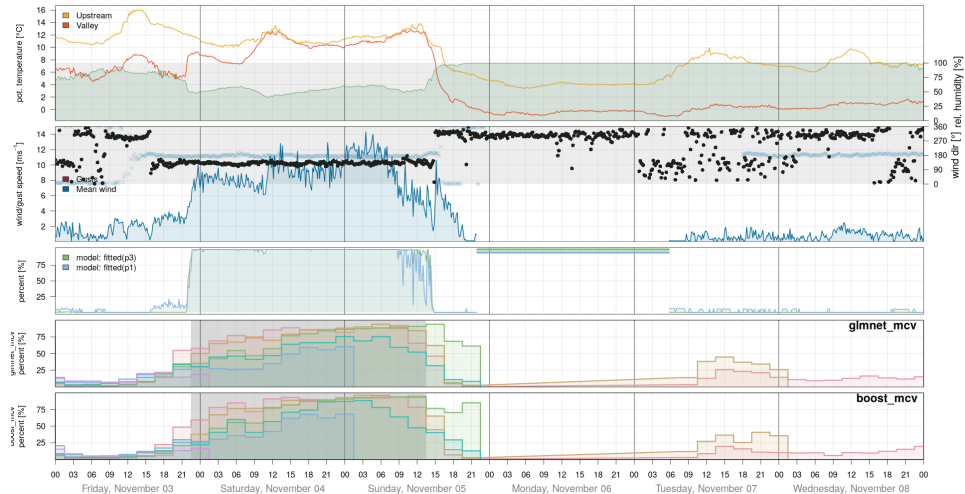
# Fallstudie: Ellbögen

Foehn Forecast Plot, Ellbogen  
2017-11-01 - 2017-11-07



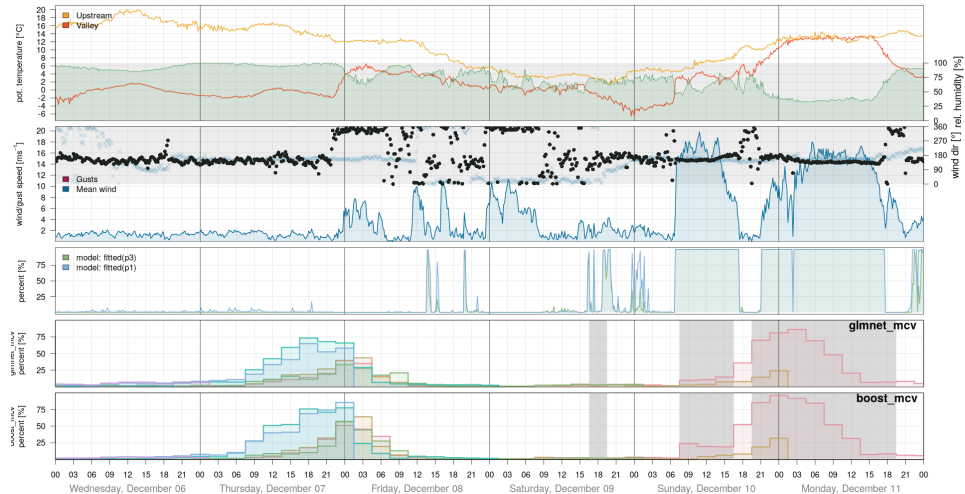
# Fallstudie: Ellbögen

Foehn Forecast Plot, Ellbogen  
2017-11-03 - 2017-11-09



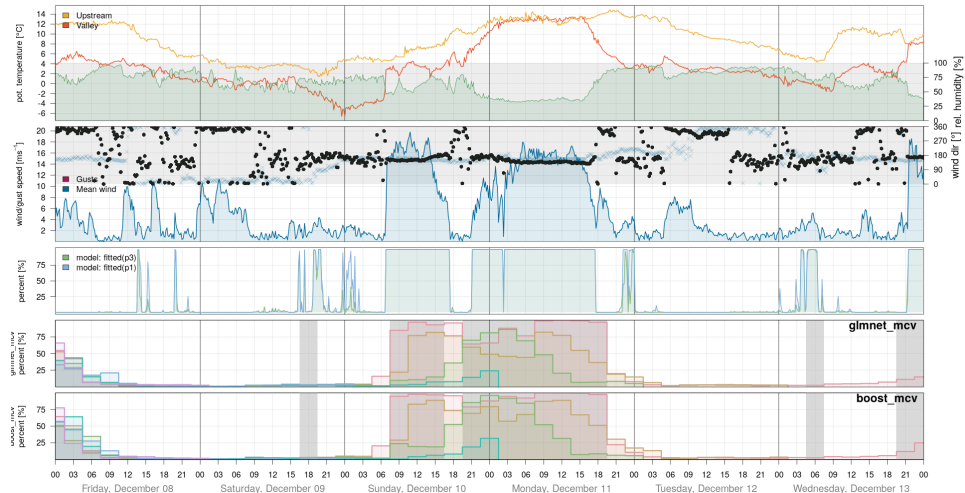
# Fallstudie: Altdorf

Föhn Forecast Plot, Altdorf  
2017-12-06 - 2017-12-12



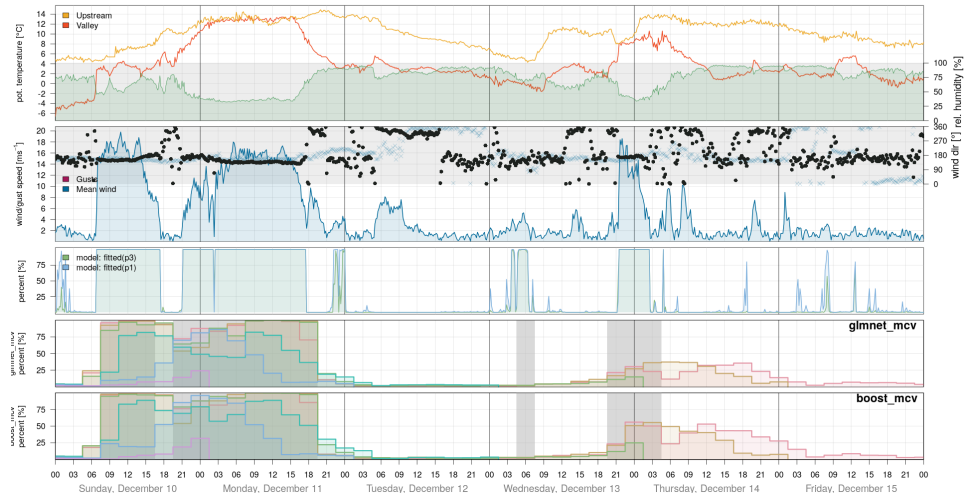
# Fallstudie: Altdorf

Föhn Forecast Plot, Altdorf  
2017-12-08 - 2017-12-14



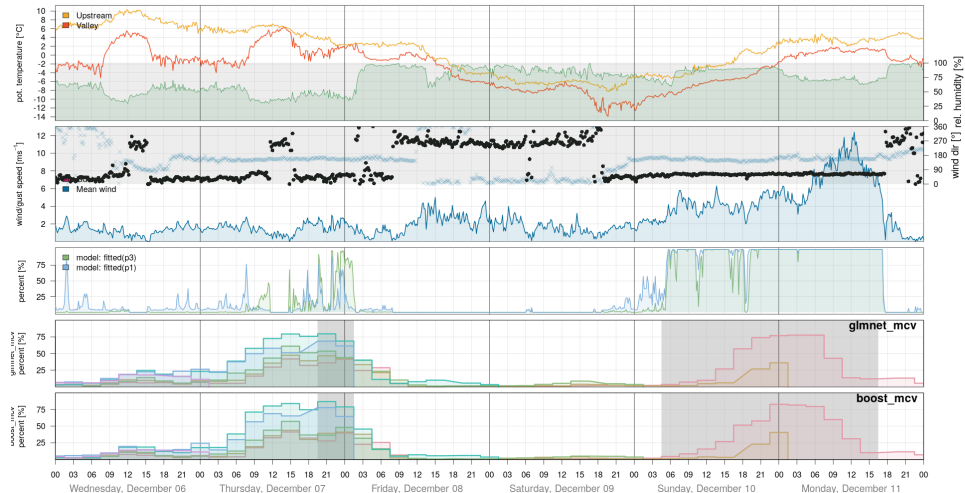
# Fallstudie: Altdorf

Föhn Forecast Plot, Altdorf  
2017-12-10 - 2017-12-16



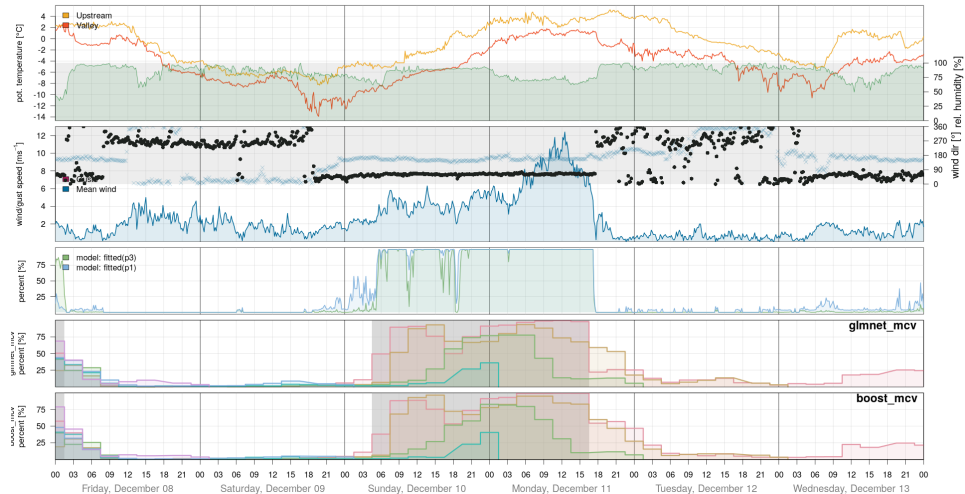
# Fallstudie: Montana

Föhn Forecast Plot, Montana  
2017-12-06 - 2017-12-12



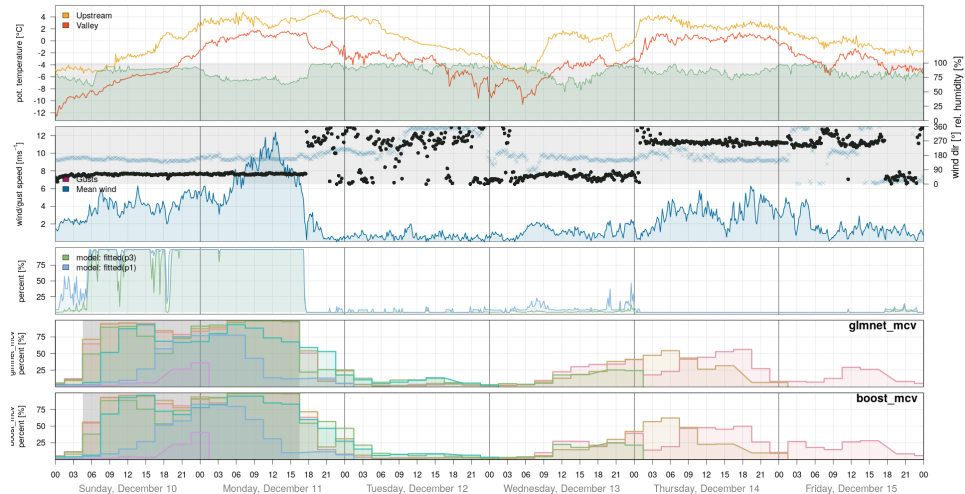
# Fallstudie: Montana

Foehn Forecast Plot, Montana  
2017-12-08 - 2017-12-14



# Fallstudie: Montana

Föhn Forecast Plot, Montana  
2017-12-10 - 2017-12-16



# Zusammenfassend

## Methodisch:

- **Vollautomatisches** Verfahren
- **Gute Performance** mit relativ simplen GLMs

# Zusammenfassend

## Methodisch:

- **Vollautomatisches** Verfahren
- **Gute Performance** mit relativ simplen GLMs
- **Gradient Boosting**  $>$  Lasso ( $N_{boost} > N_{lasso}$ )
- **Stability Selection:**
  - sehr effizient
  - führt zu schlanken Modellen mit guter Performance

# Zusammenfassend

## Methodisch:

- **Vollautomatisches** Verfahren
- **Gute Performance** mit relativ simplen GLMs
- **Gradient Boosting** > Lasso ( $N_{boost} > N_{lasso}$ )
- **Stability Selection:**
  - sehr effizient
  - führt zu schlanken Modellen mit guter Performance

## Performance:

- Skill **bis Tag 6**
- POD Stationsabhängig bis 85 % auf Tag 1

# Zusammenfassend

## Methodisch:

- **Vollautomatisches** Verfahren
- **Gute Performance** mit relativ simplen GLMs
- **Gradient Boosting** > Lasso ( $N_{boost} > N_{lasso}$ )
- **Stability Selection:**
  - sehr effizient
  - führt zu schlanken Modellen mit guter Performance

## Performance:

- Skill **bis Tag 6**
- POD Stationsabhängig bis 85 % auf Tag 1

## Erweiterungen:

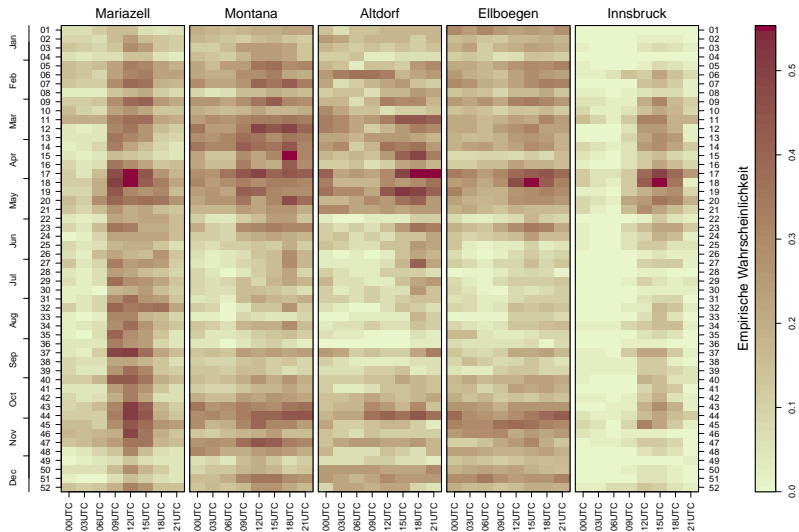
- ECMWF HRES 1-stündige Vorhersagen
- Erweiterung auf ECMWF ENS



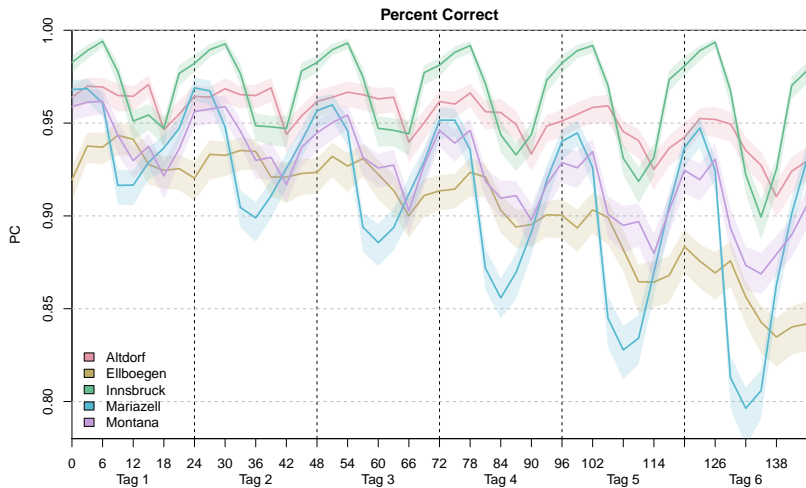
Danke für euer Interesse

[reto.stauffer@uibk.ac.at](mailto:reto.stauffer@uibk.ac.at)

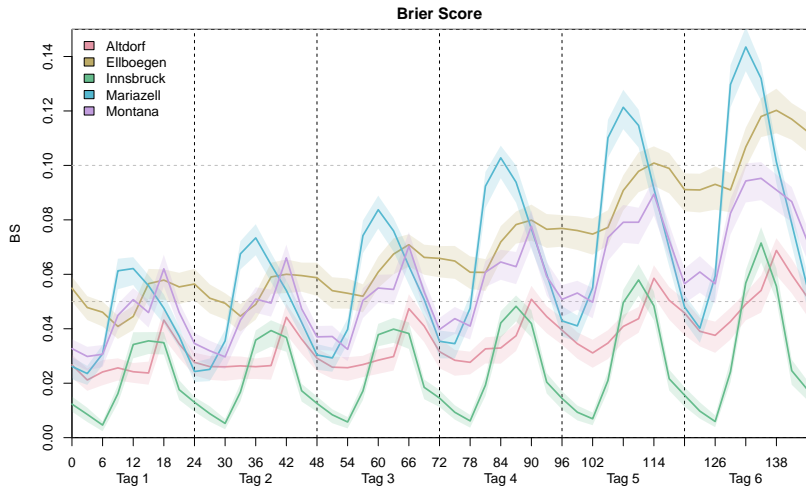
# Hovmøller Diagramme



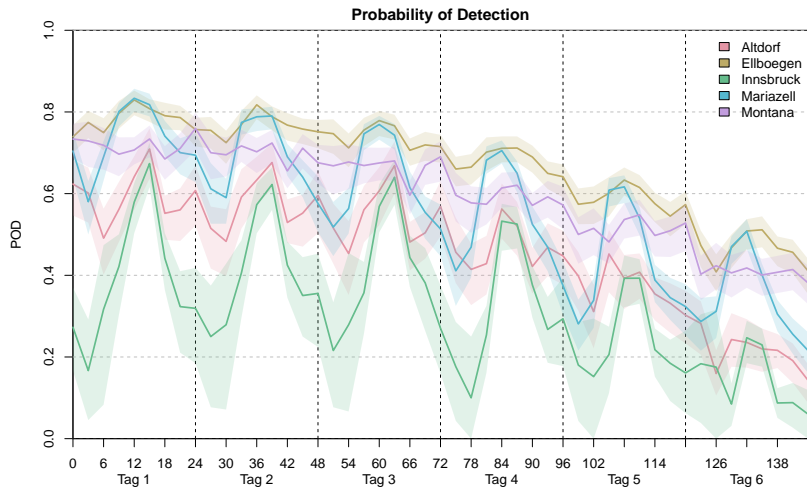
# Scores



# Scores



# Scores



# Optimierung & Variablenselektion

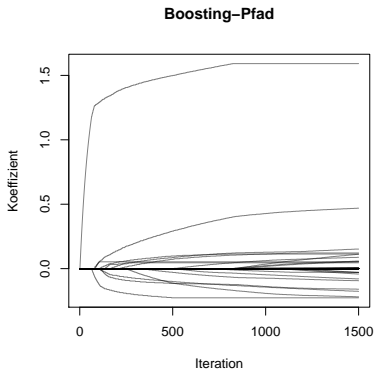
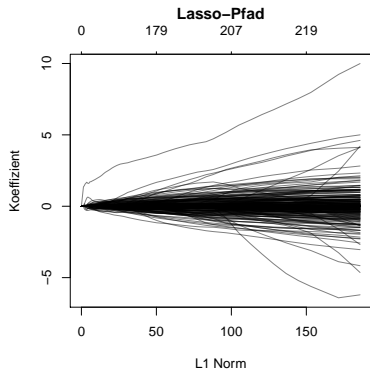
Elastic Net (Lasso):

$$\hat{\beta} = \operatorname{argmax}_{\beta \in \mathbb{R}} \left( \frac{1}{N} \sum_{i=1}^N \{y_i \cdot \eta - \log(1 + \exp(\eta))\} - \lambda \cdot \|\beta\|_1 \right) \quad (2)$$

Gradient boosting:

- 1 Start mit  $\beta^* = 0$
- 2 Berechne  $\frac{\partial}{\partial \eta} l(\beta^*)$
- 3 Update  $\beta_p^* = \beta_p^* + \beta_p * \nu$  mit stärkstem Beitrag zu  $l(\dots)$
- 4 Wiederhole 2 bis Abbruchkriterium erfüllt

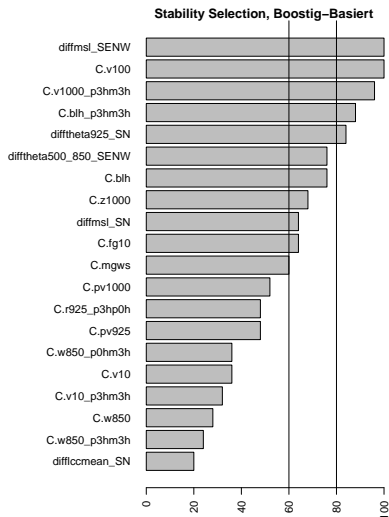
# Optimierung & Variablenselektion



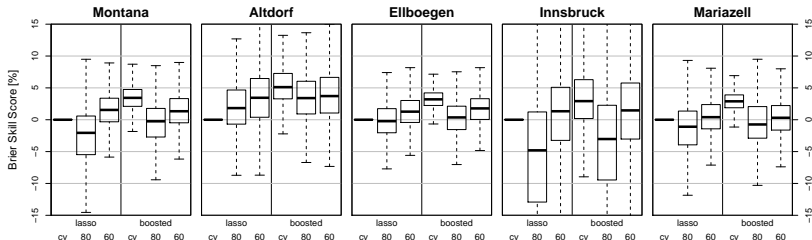
- Optimieren der Hyperparameter mittels CV
- Lasso/Boosting basierte Stability Selection

# Stability Selection

- 1 Zufälliges ziehen von 50 % des Trainingsdatensatzes
- 2 Optimierung bis  $q$  Kovariablen gewählt wurden
- 3 Merken der gewählten Kovariablen
- 4 Wiederhole Schritt 1  $B$  mal



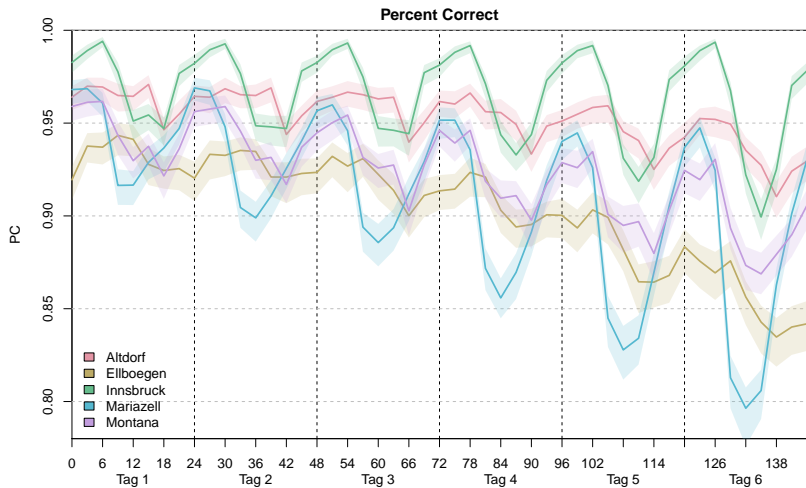
# Methodenvergleich



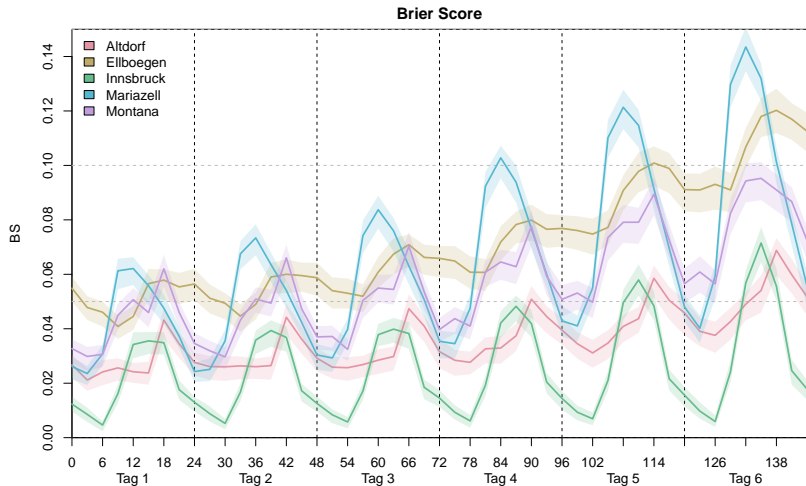
$$BS = \frac{1}{N} \sum_{i=1}^N (\pi_i - o_i)^2, \quad BSS = 1 - \frac{BS_{fcst}}{BS_{ref}} \quad (3)$$

Vergleich verschiedener Methoden (über die Jahre 2012-2017, alle Vorhersagezeitschritte).

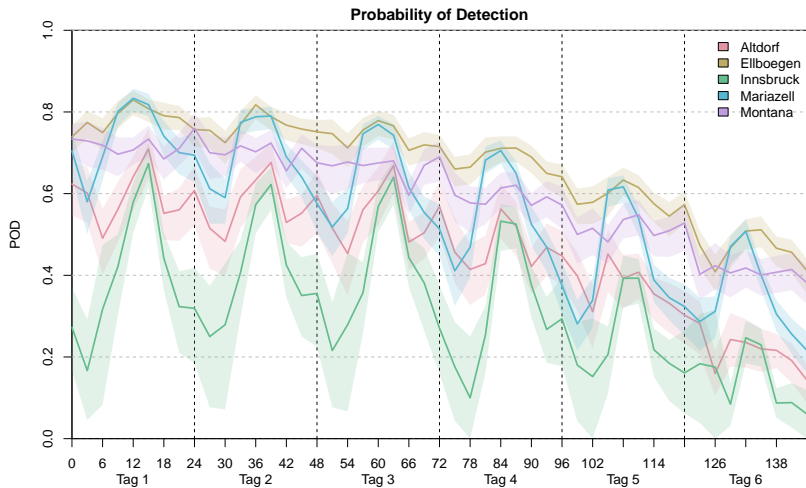
# Fehlermasse/Score Plots



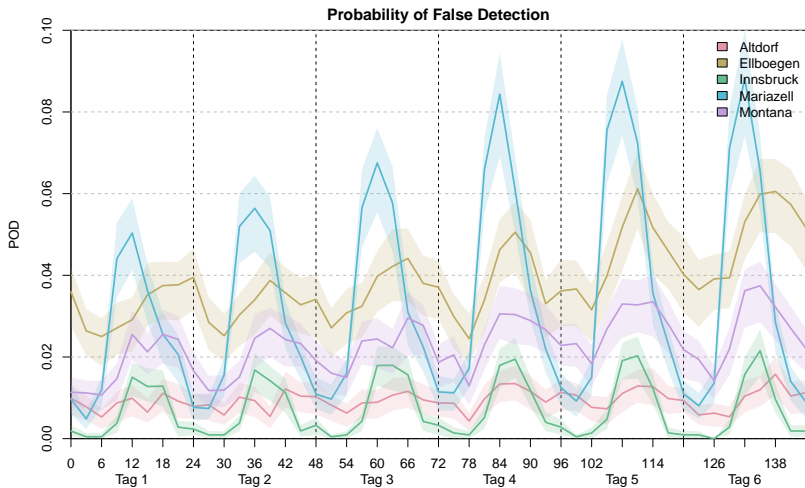
# Fehlermasse/Score Plots



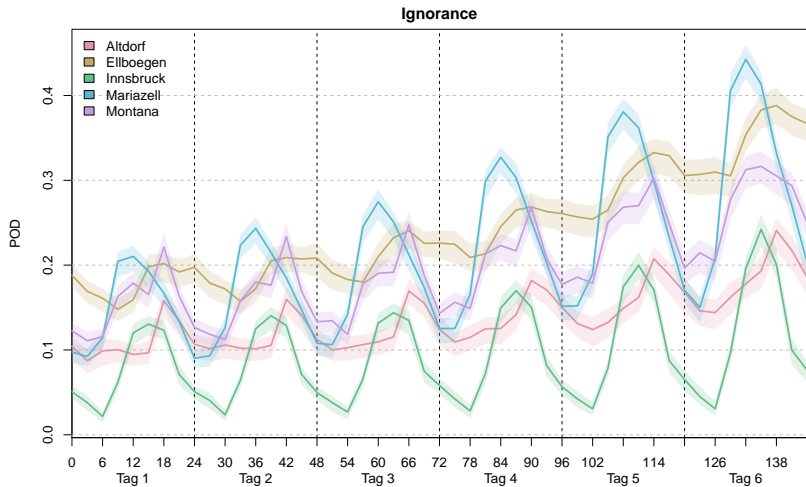
# Fehlermasse/Score Plots



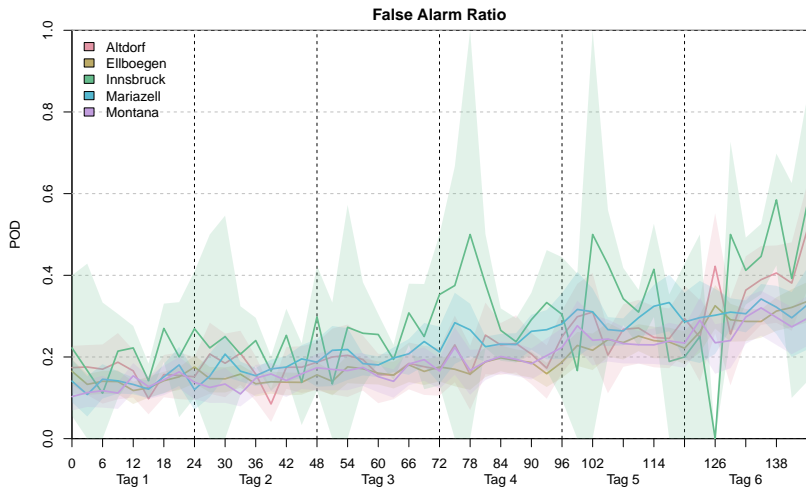
# Fehlermasse/Score Plots



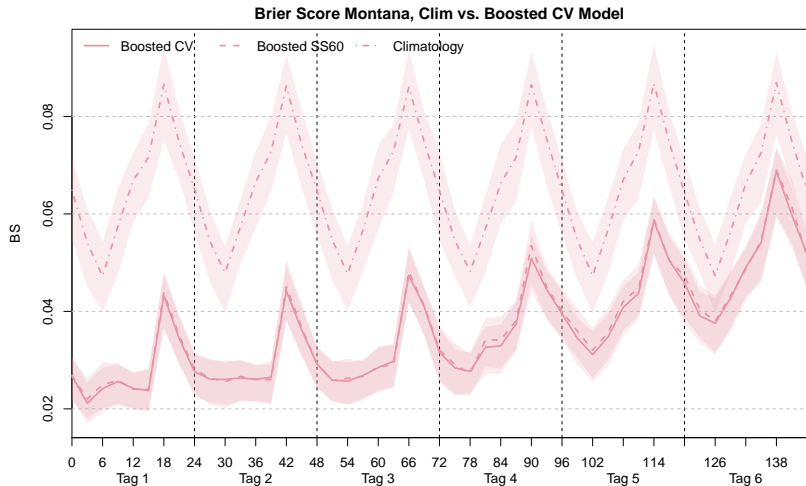
# Fehlermasse/Score Plots



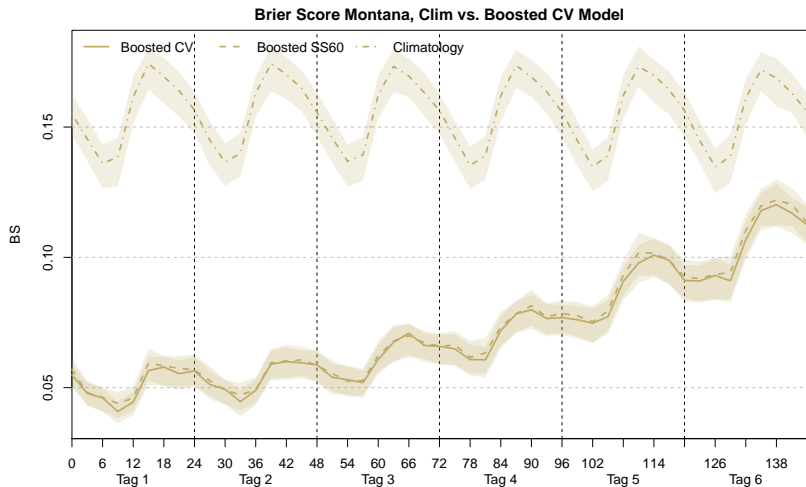
# Fehlermasse/Score Plots



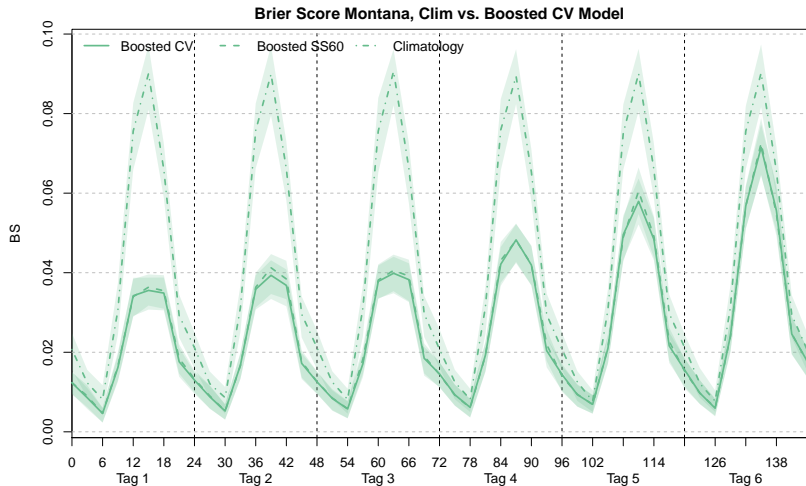
# Scores im Bezug auf Klimatologie: BS



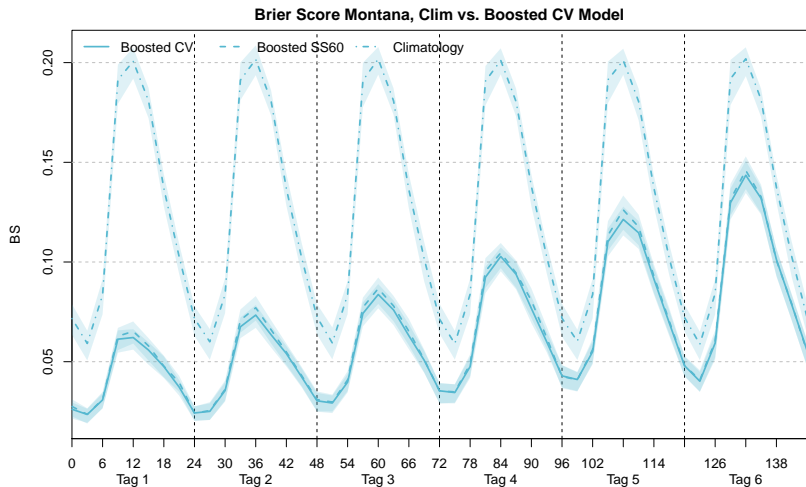
# Scores im Bezug auf Klimatologie: BS



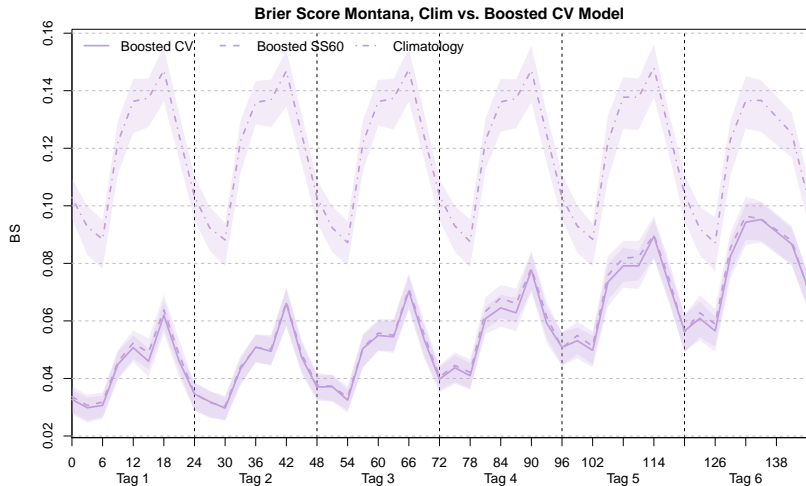
# Scores im Bezug auf Klimatologie: BS



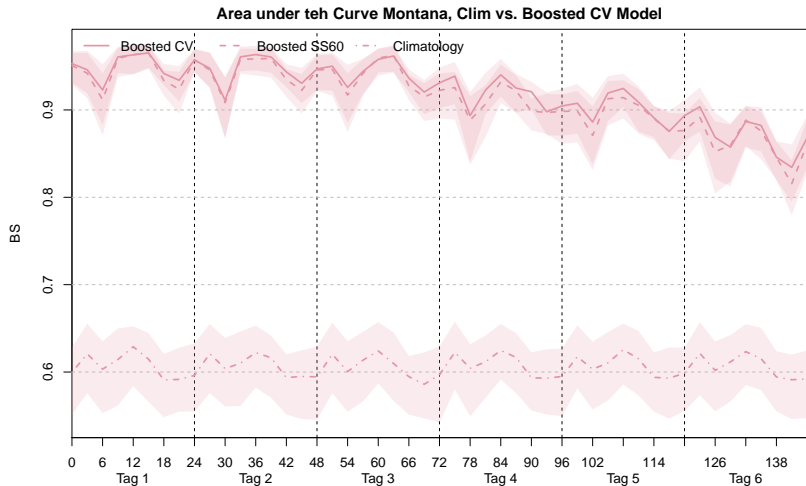
# Scores im Bezug auf Klimatologie: BS



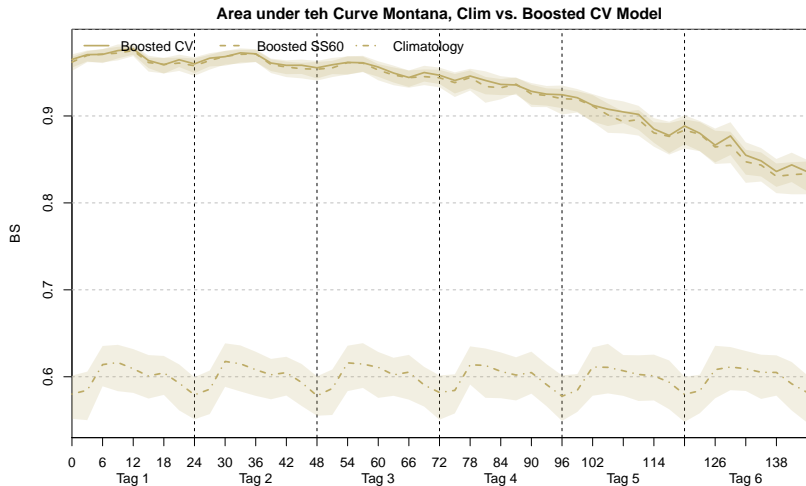
# Scores im Bezug auf Klimatologie: BS



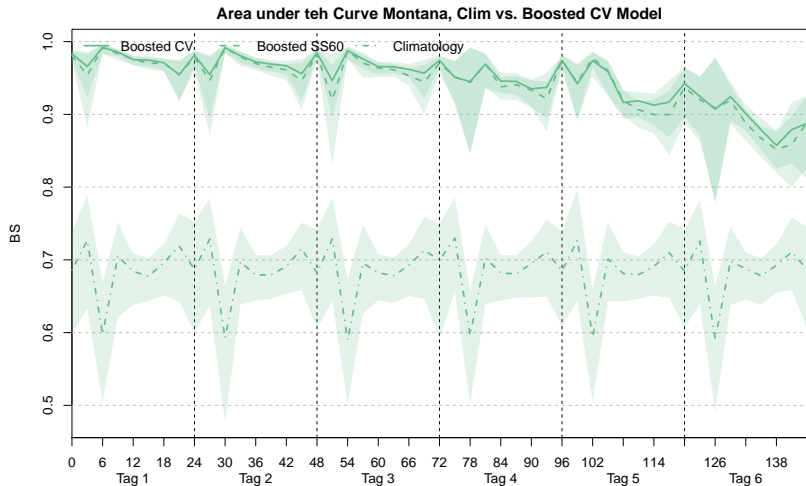
# Scores im Bezug auf Klimatologie: AUC



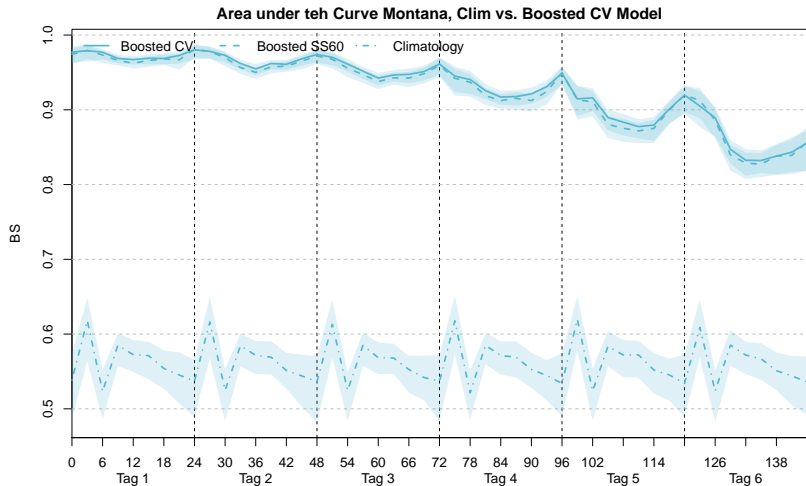
# Scores im Bezug auf Klimatologie: AUC



# Scores im Bezug auf Klimatologie: AUC



# Scores im Bezug auf Klimatologie: AUC



# Scores im Bezug auf Klimatologie: AUC

