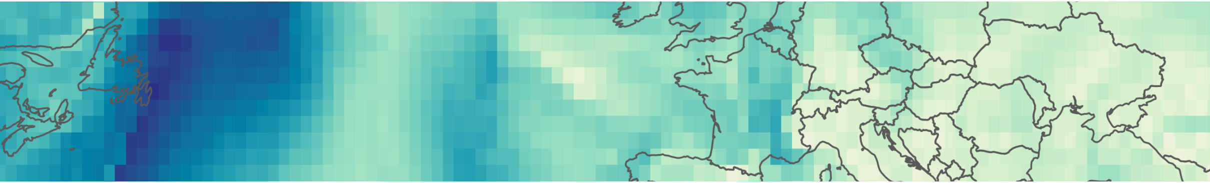
universität innsbruck



Long-term foehn reconstruction combining unsupervised and supervised learning

Reto Stauffer, Georg J. Mayr, and Achim Zeileis

www.uibk.ac.at/statistics

Foehn dynamics

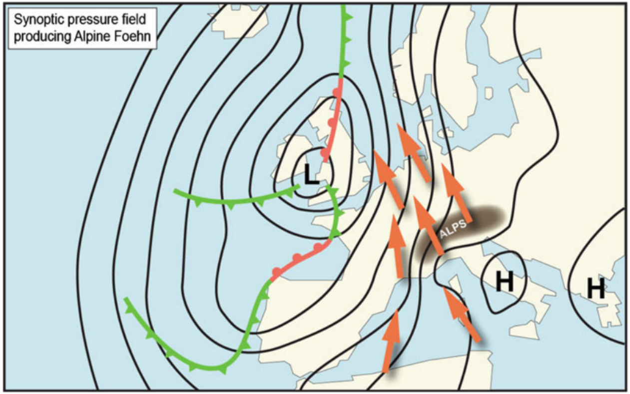
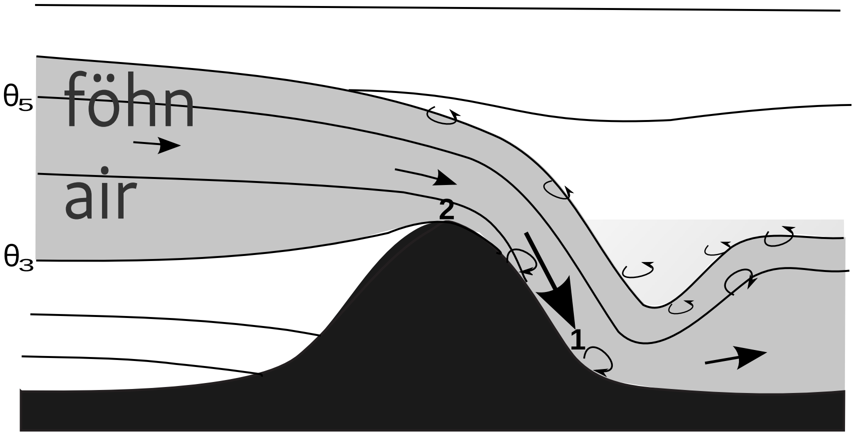


Fig. 4.4 Synoptic pressure field producing foehn in the Alps

Richner and Hächler (2013), Understanding and Forecasting Alpine Foehn.

Foehn dynamics



Provided by Georg J. Mayr (2025).

Foehn dynamics

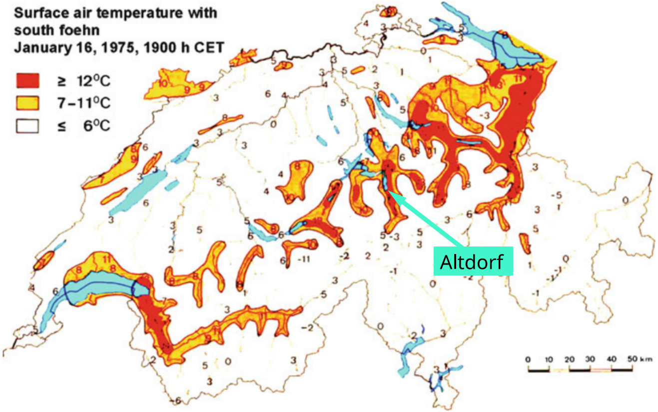
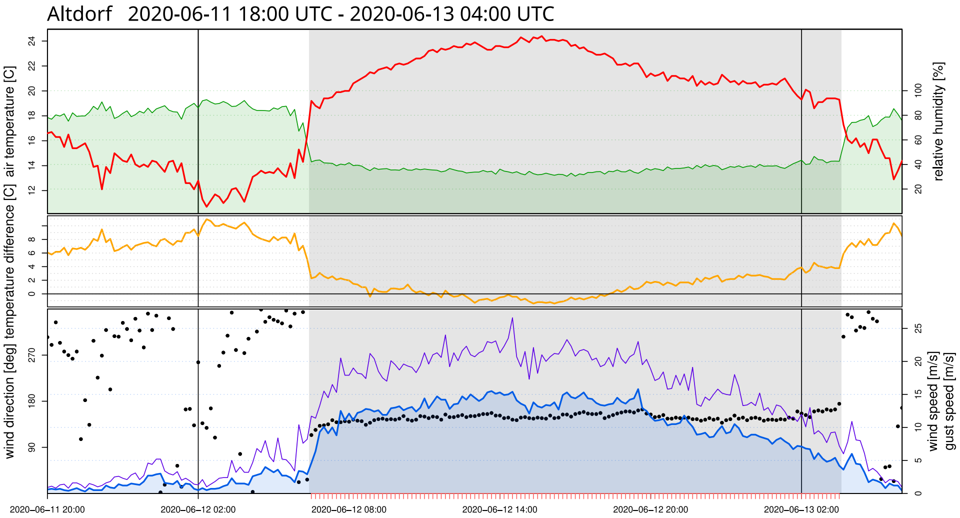


Fig. 4.22 Case study showing the main areas in Switzerland where foehn reaches the surface (Reproduced by permission of Thomas Gutermann, data kindly provided by MeteoSwiss)

Richner and Hächler (2013), Understanding and Forecasting Alpine Foehn.

Foehn dynamics



What is foehn?

Simplified

- Pressure-induced downslope wind on the **leeward** side of mountain ranges
- Airmass upstream is relatively colder than downstream, thus descends
- **Acceleration** and **alteration** (dry-adiabatic compression) on its descend
- Sudden change in humidity and temperature (often warming and drying)
- **Well-mixed** airmass between crest and downstream area

What is foehn?

Simplified

- Pressure-induced downslope wind on the **leeward** side of mountain ranges
- Airmass upstream is relatively colder than downstream, thus descends
- **Acceleration** and **alteration** (dry-adiabatic compression) on its descend
- Sudden change in humidity and temperature (often warming and drying)
- **Well-mixed** airmass between crest and downstream area

Challenge

- Cannot be observed directly
- Must be classified based on observations

Long-term foehn reconstruction

Foehn "observations"
Not directly measured.
Last 14–22 years.
Gaussian mixture model
with concomitants.

Remarks

- Based on high-resolution automated weather station (AWS) data
- Typically only available for limited periods

Long-term foehn reconstruction

Foehn "observations"
Not directly measured.
Last 14–22 years.
Gaussian mixture model
with concomitants.

Remarks

- Based on high-resolution automated weather station (AWS) data
- Typically only available for limited periods

Long-term foehn reconstruction

Foehn "observations"
Not directly measured.
Last 14–22 years.
Gaussian mixture model
with concomitants.

ERA5 reanalysis
No direct foehn information.
1940–2022.

Remarks

- Physical global atmospheric reanalysis data, hourly resolution
- Provides detailed long-term weather characterization
- No direct information about foehn

Long-term foehn reconstruction

Foehn "observations"
Not directly measured.
Last 14–22 years.
Gaussian mixture model
with concomitants.

ERA5 reanalysis
No direct foehn information.
1940–2022.

Remarks

- Physical global atmospheric reanalysis data, hourly resolution
- Provides detailed long-term weather characterization
- No direct information about foehn

Long-term foehn reconstruction

Foehn "observations"
Not directly measured.
Last 14–22 years.
Gaussian mixture model with concomitants.

ERA5 reanalysis
No direct foehn information.
1940–2022.

Supervised learning
Use foehn classification (y/n).
Covariates from ERA5.
Fit binary response model.
Reconstruct foehn 1940–2022.
Hourly temporal resolution.

Remarks

- Any binary response model can be used
- Utilized regularized logistic regression, stability selection, and XGboost

Long-term foehn reconstruction

Foehn "observations"
Not directly measured.
Last 14–22 years.
Gaussian mixture model
with concomitants.

ERA5 reanalysis
No direct foehn information.
1940–2022.

Supervised learning
Use foehn classification (y/n).
Covariates from ERA5.
Fit binary response model.
Reconstruct foehn 1940–2022.
Hourly temporal resolution.

Final result
High-res foehn probabilities.
Input for further analysis.

Remarks

- High-resolution long-term time series of foehn probabilities
- Approximately 720 000 hourly probabilities per station

Long-term foehn reconstruction

Foehn “observations”
Not directly measured.
Last 14–22 years.
Gaussian mixture model
with concomitants.

ERA5 reanalysis
No direct foehn information.
1940–2022.

Supervised learning
Use foehn classification (y/n).
Covariates from ERA5.
Fit binary response model.
Reconstruct foehn 1940–2022.
Hourly temporal resolution.

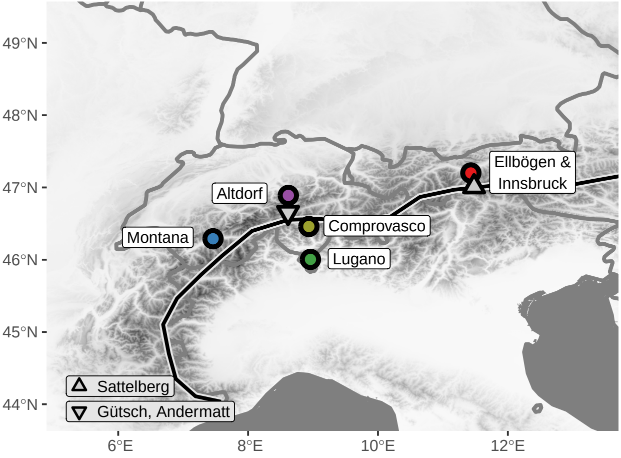
Final result
High-res foehn probabilities.
Input for further analysis.

Example application
Season-trend decomposition.
Identify long-term trends.

Remarks

- Season-trend decomposition on monthly data
- Identify possible changes/trends over the past eight decades

Study area

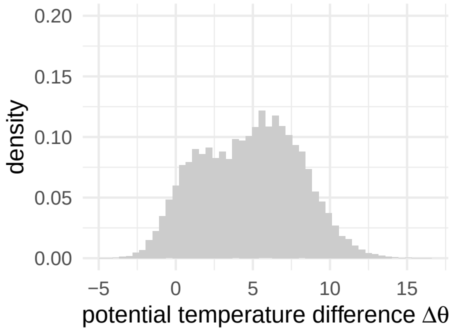


South-foehn stations
Montana, Altdorf, Ellbögen, Innsbruck (UNI)

North-foehn stations
Comprovasco, Lugano

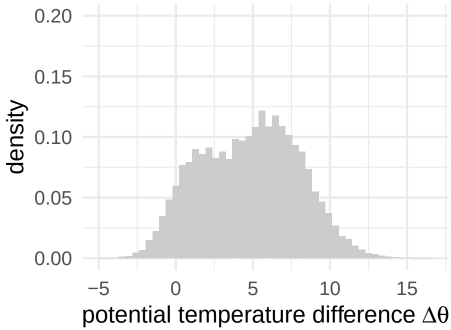
Crest stations
Gütsch and Sattelberg
for upstream information.

Foehn classification



Marginal distribution of the potential temperature difference (crest/valley), $\Delta\theta = t_{\text{crest}} + 0.01\Delta h - t_{\text{valley}}$.

Foehn classification



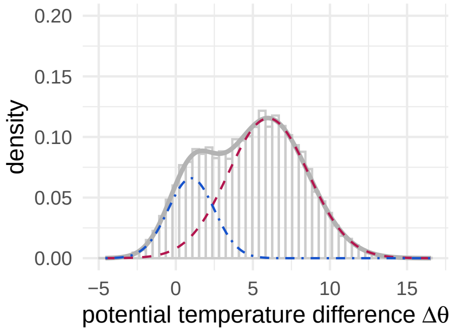
Marginal distribution of the potential temperature difference (crest/valley), $\Delta\theta = t_{\text{crest}} + 0.01\Delta h - t_{\text{valley}}$.

Assumption
Weighted combination of two Gaussian components.

Goal
Identify location and scale of both components and mixture weights.

$\Rightarrow$ Two-component Gaussian mixture model.

Foehn classification



Mixture model

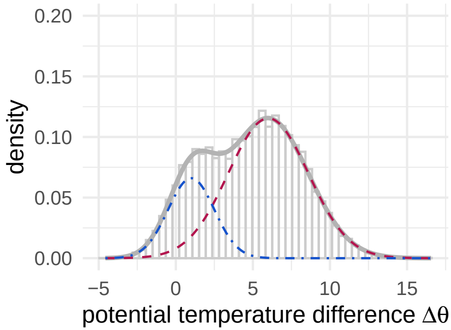
The joint density $h()$ is defined as

$$h(y, \beta) = \mathcal{N}(y, \beta_1) \cdot \pi + \mathcal{N}(y, \beta_2) \cdot (1 - \pi),$$

where the first component (blue) represents 'foehn', the second (red) 'no foehn' cases.

- y: Variable to separate main components
- β: Parameters of the main components
- π: Mixture weight

Foehn classification



Mixture model w/ concomitants

Re-define the joint density $h()$ as

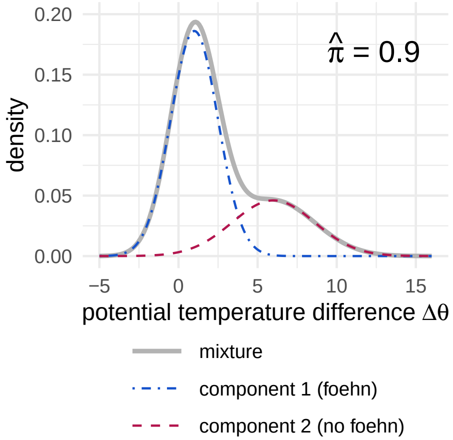
$$h(y, x, \beta, \alpha) = \mathcal{N}(y, \beta_1) \cdot \pi(x, \alpha) + \mathcal{N}(y, \beta_2) \cdot (1 - \pi(x, \alpha)),$$

with the concomitant model

$$\pi(x, \alpha) = \frac{\exp(x^T \alpha)}{1 + \exp(x^T \alpha)}.$$

- y: Variable to separate main components
- x: Concomitant variables
- β: Parameters of the main components
- α: Regression coefficients of the logit model
- π: Probability of falling into component 1

Foehn classification



Mixture model w/ concomitants

Re-define the joint density $h()$ as

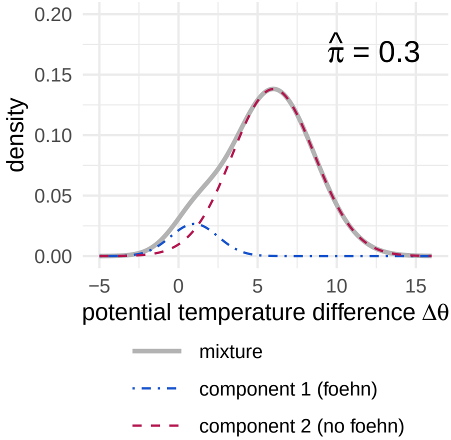
$$h(y, x, \beta, \alpha) = \mathcal{N}(y, \beta_1) \cdot \pi(x, \alpha) + \mathcal{N}(y, \beta_2) \cdot (1 - \pi(x, \alpha)),$$

with the concomitant model

$$\pi(x, \alpha) = \frac{\exp(x^T \alpha)}{1 + \exp(x^T \alpha)}.$$

- y: Variable to separate main components
- x: Concomitant variables
- β: Parameters of the main components
- α: Regression coefficients of the logit model
- π: Probability of falling into component 1

Foehn classification



Mixture model w/ concomitants

Re-define the joint density $h()$ as

$$
\begin{aligned}
h(y, x, \beta, \alpha) \quad = \quad & \mathcal{N}(y, \beta_1) \cdot \pi(x, \alpha) + \\
& \mathcal{N}(y, \beta_2) \cdot (1 - \pi(x, \alpha)),
\end{aligned}
$$

with the concomitant model

$$
\pi(x, \alpha) = \frac{\exp(x^T \alpha)}{1 + \exp(x^T \alpha)}.
$$

- y: Variable to separate main components
- x: Concomitant variables
- β: Parameters of the main components
- α: Regression coefficients of the logit model
- π: Probability of falling into component 1

Foehn classification

Estimation

The two components are separated using $y = \Delta\theta$, relative humidity and wind speed are used as covariates for the concomitant model (x). Estimate based on 14–22 years of data using the *R*-package `foehnix` (Stauffer 2023).

Foehn classification

Estimation

The two components are separated using $y = \Delta\theta$, relative humidity and wind speed are used as covariates for the concomitant model (x). Estimate based on 14–22 years of data using the *R*-package `foehnix` (Stauffer 2023).

Once the model coefficients (β, α) are estimated,
the a-posteriori probability $\text{Pr}_{\text{obs}} \in [0, 1]$ can be calculated:

$$\text{Pr}_{\text{obs}}(y, x, \beta, \alpha) = \frac{\mathcal{N}(y, \beta_1) \cdot \pi(x, \alpha)}{\mathcal{N}(y, \beta_1) \cdot \pi(x, \alpha) + \mathcal{N}(y, \beta_2) \cdot (1 - \pi(x, \alpha))}$$

Modeling foehn probability

To combine the classification with ERA5, $\mathrm{Pr}_{\mathrm{obs}}$ is upscaled to an hourly temporal resolution ($\mathrm{Pr}_{\mathrm{1h}}$). If $\mathrm{Pr}_{\mathrm{obs}} \geq 0.5$ at least half of the hour, $\mathrm{Pr}_{\mathrm{1h}} = 1$, else 0.

Supervised learning

What is needed is a binary response model of the following form:

$$\mathrm{Pr}_{\mathrm{1h}} = f_{\mathrm{learn}}(ERA5)$$

Modeling foehn probability

To combine the classification with ERA5, $\mathrm{Pr}_{\mathrm{obs}}$ is upscaled to an hourly temporal resolution ($\mathrm{Pr}_{\mathrm{1h}}$). If $\mathrm{Pr}_{\mathrm{obs}} \geq 0.5$ at least half of the hour, $\mathrm{Pr}_{\mathrm{1h}} = 1$, else 0.

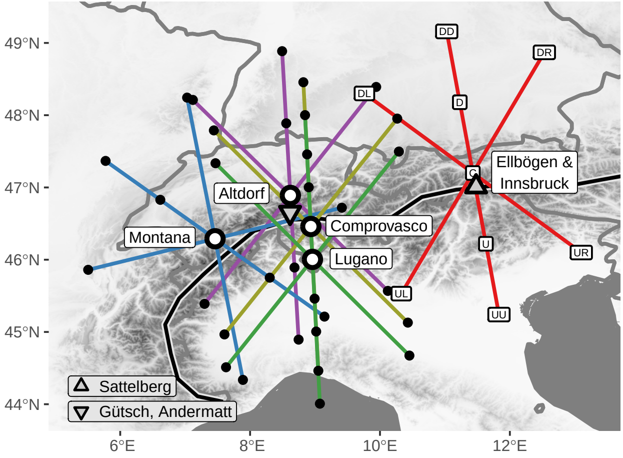
Supervised learning

What is needed is a binary response model of the following form:

$$\mathrm{Pr}_{\mathrm{1h}} = f_{\mathrm{learn}}(ERA5)$$

- **lasso:** Logistic regression with lasso (L1) regularization
- **stabsel:** Logistic regression with lasso-based stability selection
- **xgboost:** Extreme gradient boosting

ERA5 covariates



‘Direct’ set

Using 155 variables.

Only center point (C).

‘Local’ derived variables.

‘Full’ set

Additional spatial, temporal, and spatio-temporal derived variables to incorporate large-scale information (497 in total).

Long-term foehn reconstruction

Models estimated

- Separate models for each hour of the day to account for the time of day
- In total 144 models per station (24 hours, 3 models, 2 covariate sets)

Long-term foehn reconstruction

Models estimated

- Separate models for each hour of the day to account for the time of day
- In total 144 models per station (24 hours, 3 models, 2 covariate sets)

Main result

Applied to ERA5 these models result time series of $\hat{\Pr}_{1h}$ 1940–2022 (83 years).

Long-term foehn reconstruction

Models estimated

- Separate models for each hour of the day to account for the time of day
- In total 144 models per station (24 hours, 3 models, 2 covariate sets)

Main result

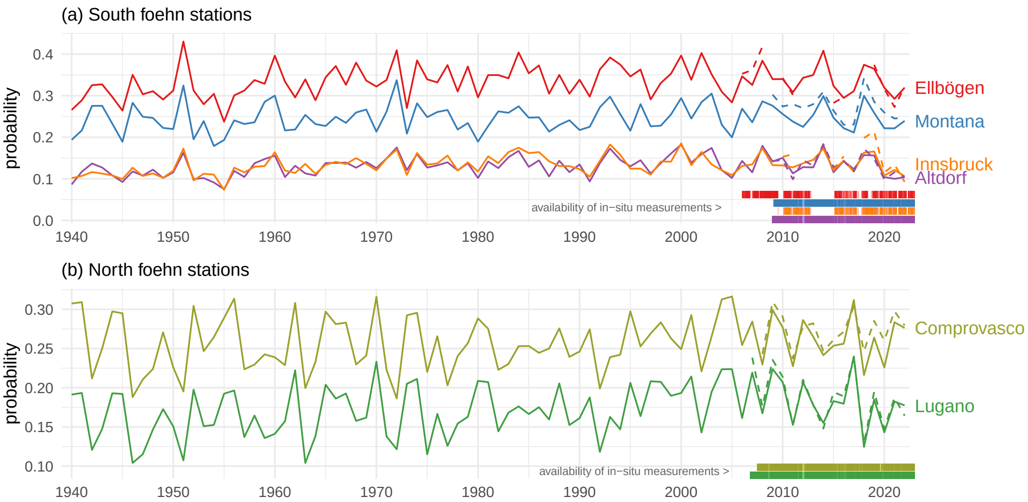
Applied to ERA5 these models result time series of $\hat{\Pr}_{1h}$ 1940–2022 (83 years).

Novelty

Foehn time series with such a high temporal resolution for such a long period did not exist yet. This allows for new interesting research questions.

Long-term foehn reconstruction

Long-term foehn reconstruction



Long-term foehn reconstruction

Altdorf: Unique oportunity to compare our reconstruction with the ‘long foehn time series’ (Gutermann *et al.* 2012, Richner *et al.* 2014).

Long-term foehn reconstruction

Altdorf: Unique oportunity to compare our reconstruction with the ‘long foehn time series’ (Gutermann *et al.* 2012, Richner *et al.* 2014).

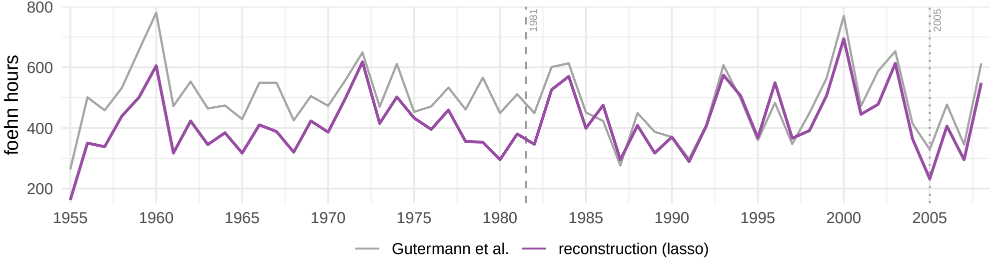


Figure: Comparison of the annual number of foehn hours at Altdorf 1955–2008. Gutermann based on observations, high-resolution AWS measurements starting in June 1981. Note that our reconstruction has not seen any data before 2005.

Model performance

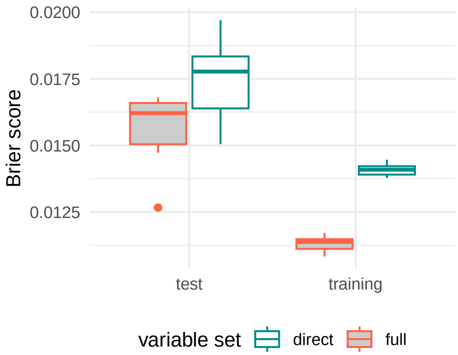


Figure: Brier scores to compare the benefit of the ‘full’ covariate set (left; lasso model) and the performance of the different models (right) for Altdorf. Lower scores are better. Based on a six-fold cross-validation using 12 years of data.

Model performance

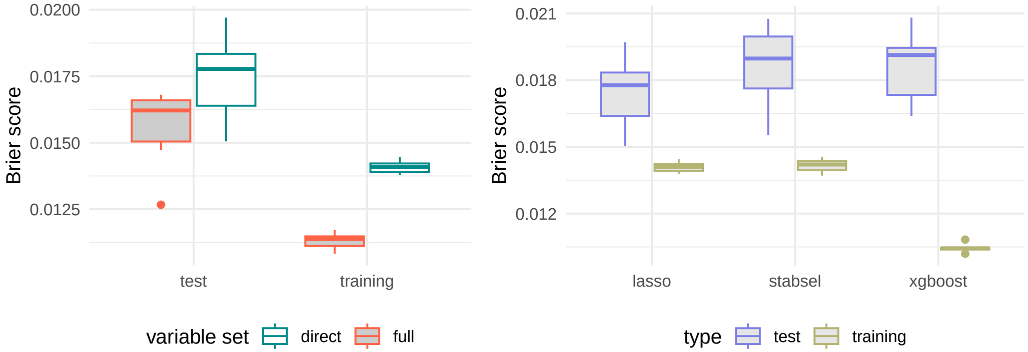
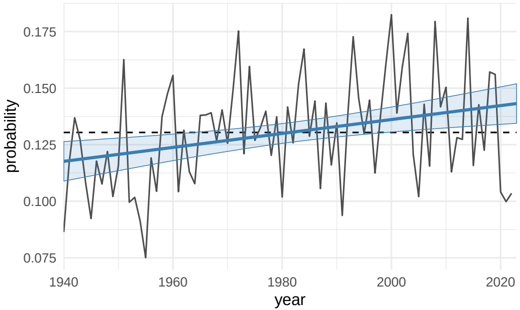


Figure: Brier scores to compare the benefit of the ‘full’ covariate set (left; lasso model) and the performance of the different models (right) for Altdorf. Lower scores are better. Based on a six-fold cross-validation using 12 years of data.

Season-trend decomposition

Annual mean foehn probability with
smooth trend estimate (and confidence interval)



Monthly mean foehn probability
per decade with seasonal effect

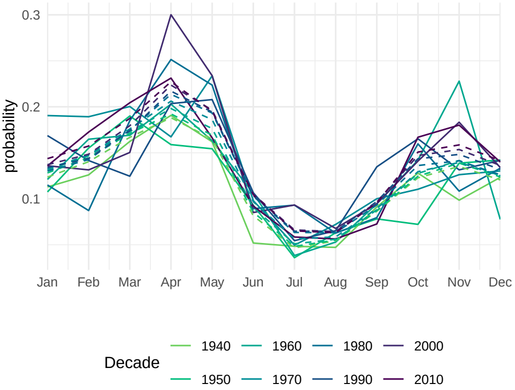


Figure: Smooth long-term trend (left) and smoothly varying seasonal changes (right) on monthly means of daily maximum for Altdorf (*R*-package `stR`, Dokumentov *et al.* 2023).

Key takeaways and future directions

What we learned

- The novel combination of unsupervised and supervised statistical learning performs well, en par (or even better) than previously published methods

Key takeaways and future directions

What we learned

- The novel combination of unsupervised and supervised statistical learning performs well, en par (or even better) than previously published methods
- Reconstructions are a powerful tool to get more insights into the past, allowing to produce largely homogeneous results

Key takeaways and future directions

What we learned

- The novel combination of unsupervised and supervised statistical learning performs well, en par (or even better) than previously published methods
- Reconstructions are a powerful tool to get more insights into the past, allowing to produce largely homogeneous results

Model performance & large-scale information

- Large scale information from the surrounding is strongly beneficial (foehn)
- All binary response models performed similarly–the simplest (‘lasso’) emerged as the overall most efficient

Key takeaways and future directions

What we learned

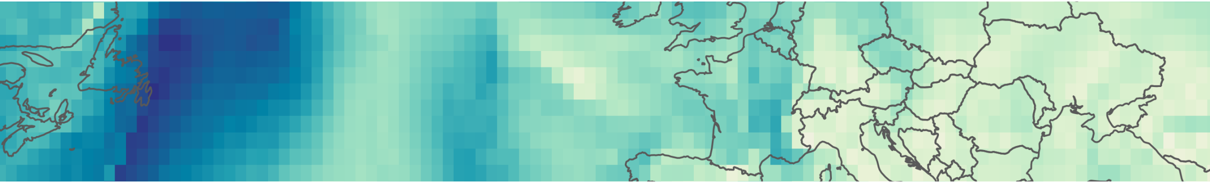
- The novel combination of unsupervised and supervised statistical learning performs well, en par (or even better) than previously published methods
- Reconstructions are a powerful tool to get more insights into the past, allowing to produce largely homogeneous results

Model performance & large-scale information

- Large scale information from the surrounding is strongly beneficial (foehn)
- All binary response models performed similarly–the simplest (‘lasso’) emerged as the overall most efficient

What’s next

- Adjust and apply ‘long-term foehn reconstruction’ to stations in California
- Apply to other target variables (lightning) and leverage the reconstruction to identify inconsistencies in existing datasets



Thank you for your attention,
looking forward to the upcoming talks!

References

Stauffer R, Zeileis A, Mayr GJ (2024). “Long-Term Foehn Reconstruction Combining Unsupervised and Supervised Learning.” *International Journal of Climatology*, **44**(16), 5890–5901, `doi:10.1002/joc.8673`

Stauffer R (2023). “foehnix: A Toolbox for Automated Foehn Classification based on Mixture Models.” *R package version 0.1-6*, `https://retostauffer.github.io/Rfoehnix/`

Richner H, Dürr B, Gutermann T, Bader S (2014). “The use of automatic station data for continuing the long time series (1864 to 2008) of foehn in Altdorf.” *Meteorologische Zeitschrift*, 23, 159–166.

Gutermann T, Dürr B, Richner H, Bader S (2012). “Föhnklimatologie Altdorf: Die lange Reihe (1864–2008) und ihre Weiterführung, Vergleich mit anderen Stationen.”, *Fachbericht MeteoSchweiz*, 241, 1–53, https://www.meteoswiss.admin.ch/...

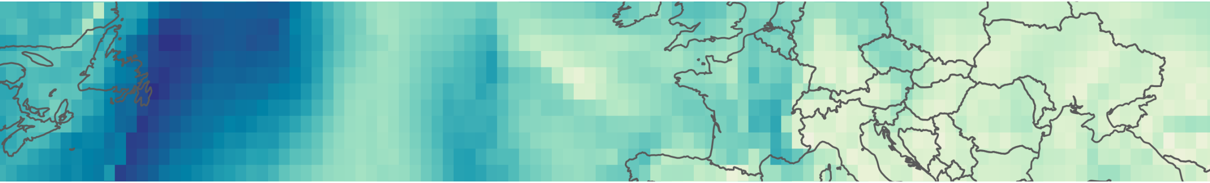
Dokumentov A, Hyndman RJ (2023). “STR: Seasonal-Trend Decomposition Using Regression.” *INFORMS Journal on Data Science*, 1(1), 50–62, `doi:10.1287/ijds.2021.0004`

Meinshausen N, Bühlmann Peter (2010). “Stability Selection.” *Journal of the Royal Statistical Society B*, 72(4), 417–473, `doi:10.1111/j.1467-9868.2010.00740.x`

Friedman JH, Hastie T, Tibshirani R (2010). “Regularization Paths for Generalized Linear Models via Coordinate Descent.” *Journal of Statistical Software*, 33(1), 1–22, `doi:10.18637/jss.v033.i01`

Tay JK, Narasimhan B, Hastie Trevor (2023), “Elastic Net Regularization Paths for All Generalized Linear Models.” *Journal of Statistical Software*, 106(1), 1–31, `doi:10.18637/jss.v106.i01`

Chen Tianqi, Guestrin C (2016), “XGBoost: A Scalable Tree Boosting System.” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794, `doi:10.1145/2939672.2939785`



Appendix and backup slides

Observations

	Type	Location	Data availability
△ Gütsch (Andermatt)[1]	crest	46.653N/8.616E 2286m	2005-01-01–2023-12-31 (95.3%)
Altdorf[1]	south	46.890N/8.620E 438m	2005-01-01–2022-12-31 (78.1%)
Montana[1]	south	46.290N/7.460E 1423m	2005-01-01–2022-12-31 (77.1%)
Comprovasco[1]	north	46.460N/8.935E 576m	2005-01-01–2022-12-31 (85.8%)
Lugano[1]	north	46.004N/8.960E 205m	2005-01-01–2022-12-31 (90.2%)
△ Sattelberg[2]	crest	47.011N/11.479E 2107m	2006-01-01–2022-12-31 (75.0%)
Ellbögen[2]	south	47.200N/11.430E 1080m	2006-01-01–2022-12-31 (92.0%)
. (Universität) Innsbruck[3]	south	47.260N/11.385E 578m	2009-06-21–2022-12-30 (99.1%)

Table: Station type, location and data availability; begin and end date plus percent available within period. Observations provided by the Swiss national weather service (1; MeteoSwiss), the University of Innsbruck (2) and the Austrian national weather service (3; GeoSphere Austria). Four stations are used to model south foehn, two to model north foehn and two serving information at the mountain crest (△; cf. Type).

ERA5

CDS ERA5 fields name	
100m_u_component_of_wind	100m_v_component_of_wind
10m_u_component_of_wind	10m_v_component_of_wind
10m_wind_gust_since_previous_post_processing	instantaneous_10m_wind_gust
2m_dewpoint_temperature	2m_temperature
surface_pressure	mean_sea_level_pressure
eastward_gravity_wave_surface_stress	northward_gravity_wave_surface_stress
convective_precipitation	large_scale_precipitation
total_precipitation	boundary_layer_dissipation
boundary_layer_height	charnock
friction_velocity	gravity_wave_dissipation
high_cloud_cover	low_cloud_cover
medium_cloud_cover	total_cloud_cover
total_column_cloud_liquid_water	surface_net_solar_radiation
surface_net_thermal_radiation	surface_sensible_heat_flux
surface_solar_radiation_downwards	surface_thermal_radiation_downwards

Table: List of ERA5 single level field retrieved.

ERA5

CDS ERA5 fields name	
divergence	geopotential
potential_vorticity	specific_humidity
temperature	u_component_of_wind
v_component_of_wind	vertical_velocity
specific_cloud_liquid_water_content	vorticity

Table: List of ERA5 pressure level fields retrieved; each on the standard pressure levels 500 hPa, 700 hPa, 750 hPa, 800 hPa, 850 hPa and 900 hPa.

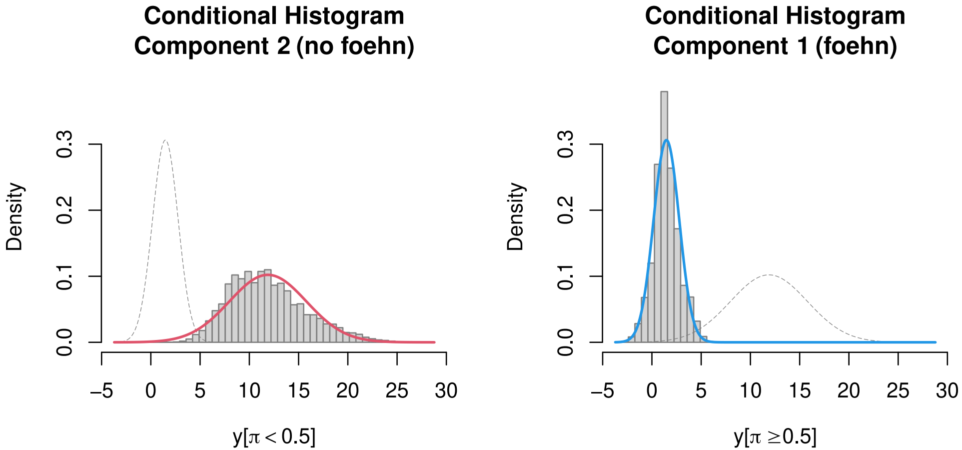
ERA5

The ‘full’ variable set contains a series of spatial and spatio-temporal information such as:

- surface pressure differences between `C` to `U`, `U` to `C`, `UL` to `DR` and others,
- potential temperature differences between `C` to `U`, `U` to `C`, `UL` to `DR` and others,
- differences in vertical temperature differences between `C` to `U`, `U` to `C`, `UL` to `DR` and others,
- sum of precipitation on the upwind side (sum of `U`, `UU`, `UR`, `UL`) as well as on the downwind side (sum of `D`, `DD`, `DR`, `DL`) as well and the difference between these two sums,
- mean cloud cover on the upwind side (sum of `U`, `UU`, `UR`, `UL`) as well as on the downwind side (sum of `D`, `DD`, `DR`, `DL`) as well and the difference between these two means.

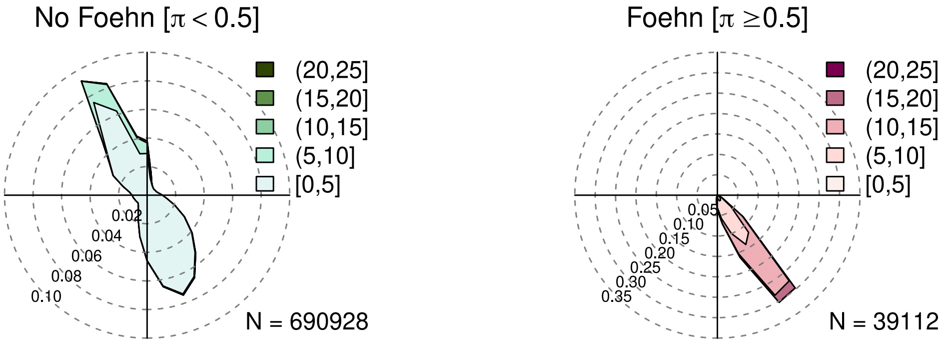
Foehn classification

Result of estimated two-component mixture model for Altdorf (CH).



Foehn classification

Result of estimated two-component mixture model for Altdorf (CH).



Foehn classification

Temporal upscaling

To be combined with ERA5 data, $\mathrm{Pr}_{\mathrm{obs}}$ is upscaled to an hourly temporal resolution using:

$$\mathrm{Pr}_{1\mathrm{h}} = \begin{cases} \text{missing} & \text{if} \quad \sum \mathrm{Pr}_{\mathrm{obs}} \in [0-1] < 4 \\ \text{foehn} & \text{if} \quad \frac{1}{N} \sum \left(\mathrm{Pr}_{\mathrm{obs}} \geq 0.5\right) \geq 0.5 \\ \text{no foehn} & \text{else} \end{cases}$$

An hour is considered a foehn event ($\mathrm{Pr}_{1\mathrm{h}}$) if classified at least four times per hour, and 50% or more are classified as ‘foehn’.

Binary response models: Lasso

Due to the large number of possible covariates from ERA5 and the fact that many of these variables are highly correlated, penalization is required. This study employs lasso (least absolute shrinkage and selection operator) with L1 regularization, optimizing the following penalized log-likelihood:

$$(\hat{\beta}_0, \hat{\beta}) = \underset{(\beta_0,\beta)\in\mathcal{R}^k}{\text{argmin}} \left[\frac{1}{2N} \sum_{i=1}^{N} (\text{Pr}_{\text{obs},i} - \beta_0 - \text{ERA5}^\top \beta)^2 + \lambda ||\beta||_1 \right],$$

with $\text{ERA5} \in \mathcal{R}^{N\times k}$, where N is the number of hourly foehn probabilities (Sec. 3.1) and k the number of available covariates (Sec. 2.2).
To find the optimal tuning parameter $\hat{\lambda}$, a 30-fold cross-validation is performed which yields the optimal (regularized) set of regression coefficients $\hat{\beta}_0, \hat{\beta}$ using the *R* package `glmnet` (with `s = "lambda.min"`).

Binary response models: Stability selection

The penalized regression model (lasso) is estimated $M = 200$ times with $\lambda \in [0, \infty]$. The first $K = 40$ parameters which enter the model ($\neq 0$) will be memorized. Covariates which have been selected more than 60% of the time (at least 121/200 times) are then used to estimate an additional unregularized logistic regression model.

1. Whilst iteration is ≤ 200:
    1. Draw stratified subsample of size $N/2$
    2. Estimate regularized coefficients $(\hat{\beta}_0, \hat{\beta})$ for different λs
    3. Extract and keep the name of the first $K = 40$ covariates entering the model
2. Select the covariates which a selection frequency > 0.6
3. Estimate final (unregularized) logistic regression model with the covariates from Step 2

Allows for automatic variable selection, strongly reducing the complexity of the final model without losing much predictive performance.

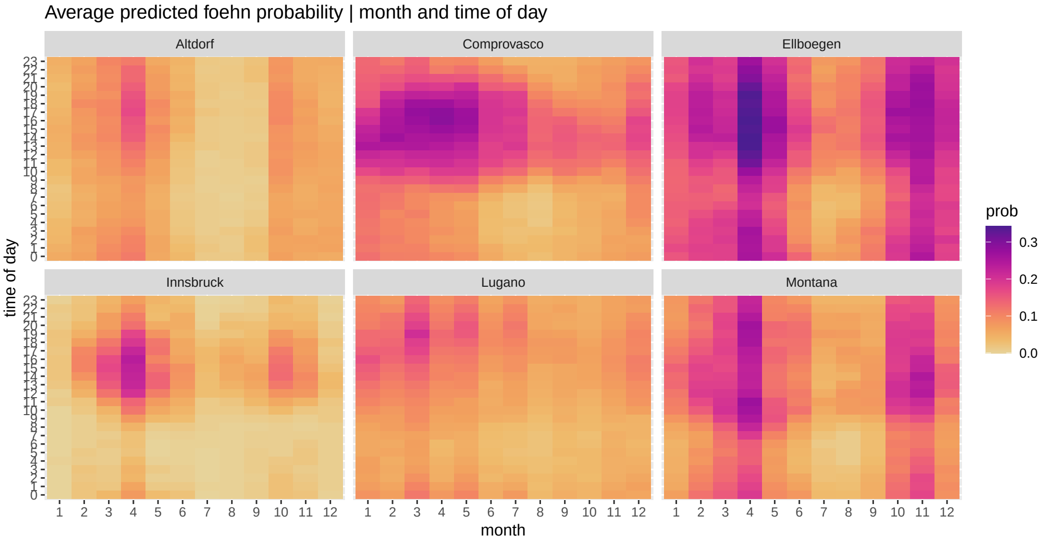
Binary response models: XGboost

We performed a (limited) grid search using the *R* package `xgboost` with the following parameter space:

- `eta`, learning rate: $\{0.1, 0.125, 0.15, 0.2\}$
- `max_depth`, maximum depth of each weak learner (tree): $\{5, 10, 20\}$
- `min_child_weight`, minimum sum of instance weight needed in a child: $\{2, 4, 6\}$
- `gamma`, minimum loss reduction required: $\{2, 5, 10\}$

This yields an overall number of 108 unique combinations. For each combination, a 10-fold cross-validation using random stratified subsamples and a maximum of 20 boosting iterations is performed. It is then evaluated which parameter set and boosting iteration resulting in the lowest average out-of-sample error to estimate one final model using all data.

Diurnal variability



Season-trend decomposition

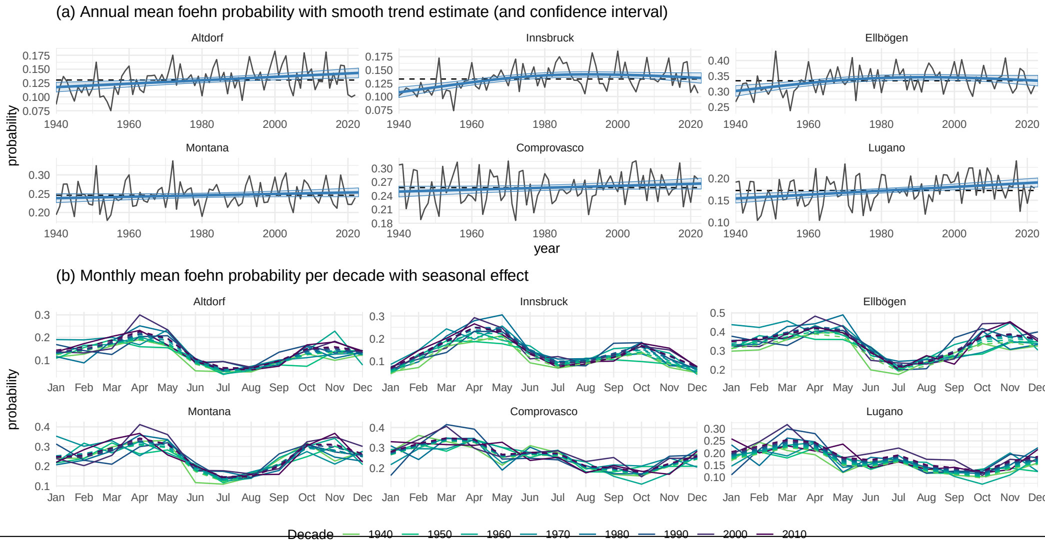
Did the occurrence of foehn change with the changing climate?

Separate long-term signal from the (large) year-by-year and within-year variability using a regression based season-trend decomposition model of the following form (Dokumentov 2023):

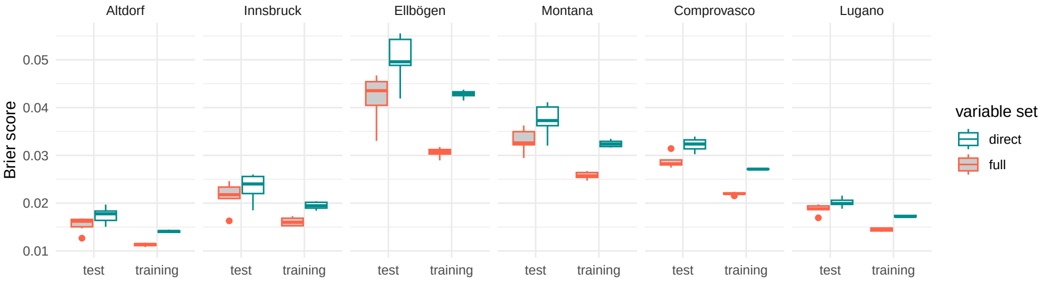
$$y_t = T_t + \sum_{m=1}^{12} S_t^{(m)} + R_t \text{ with } t \in 1, \ldots, J.$$

- y_t: monthly mean of the daily maxima of the reconstruction $\hat{\text{Pr}}_{\text{1h}}$
- T_t: smooth long-term trend
- $S_t^{(m)}$: smoothly changing seasonal component
- R_t: remainder (or residual)
- $m \in 1, \ldots, 12$: months
- J: the sample size; $83\text{years} \times 12 = 996$

Season-trend decomposition



Benefit of full covariate set



Comparison of supervised learners