https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022

Atmospheric Measurement Techniques Discussions

Prediction of Alpine Foehn from time series of GNSS troposphere products using machine learning

Matthias Aichinger-Rosenberger[1], Elmar Brockmann[2], Laura Crocetti[1], Benedikt Soja[1], and Gregor Moeller[1]

[1]Institute of Geodesy and Photogrammetry, ETH Zürich, 8093 Zurich, Switzerland
[2]Swiss Federal Office of Topography (swisstopo), 3084 Wabern bei Bern, Switzerland

Correspondence: Matthias Aichinger-Rosenberger, maichinger@ethz.ch

Abstract. Remote sensing of water vapor using the Global Navigation Satellite System (GNSS) is a well-established technique and reliable data source for Numerical Weather Prediction (NWP). One of the phenomena rarely studied using GNSS are foehn winds. Since foehn winds are associated with significant humidity gradients between lee/luv side of a mountain range, tropospheric estimates from GNSS are also affected by their occurrence. Time series reveal characteristic features like distinctive minima/maxima and significant decrease in correlation between the stations. However, detecting such signals becomes increasingly difficult for large data sets. Therefore, we suggest the application of machine learning algorithms for detection and prediction of foehn events from GNSS troposphere products. The present study uses long-term time series of high-quality GNSS troposphere products from the Automated GNSS Network Switzerland (AGNES) as well as records of operational foehn index to investigate the performance of several different classification algorithms based on appropriate statistical metrics. The two best-performing algorithms are fine-tuned and employed on two years of test data. The results show very promising results, especially when reprocessed GNSS products are utilized. Detection- and alarm-based measures reach levels of 70-85% for both tested algorithms and thus are comparable to those from studies using data from meteorological stations and NWP. For operational prediction, some limitations due to the availability and quality of GNSS products in near-real time (NRT) exist. However, they might be mitigated to a significant extend by provision of additional NRT products and improved data processing in the future.

1 Introduction

Global Navigation Satellite Systems (GNSS) are used extensively for positioning and navigation applications worldwide. Additionally, they enable users to retrieve information about the state of the Earth's atmosphere, in particular the distribution of water vapor. This technique, commonly referred to as GNSS Meteorology, was first proposed three decades ago (Bevis et al., 1992) and is still gaining increasing interest from the scientific community as well as meteorological institutions. The retrieval of atmospheric information from GNSS is based on the fact that electromagnetic signals (such as GNSS signals) are delayed when traveling through specific layers of the atmosphere. The delay experienced by a GNSS signal in the lowest part of the atmosphere (troposphere) is proportional to the water vapor content along the signal path. This fact is typically exploited in

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022


Atmospheric Measurement Techniques Discussions

GNSS Meteorology by introducing GNSS-derived atmospheric parameters like the Zenith Wet Delay (ZWD) or the Zenith Total Delay (ZTD) in data assimilation schemes. In numerous studies, a positive impact has been demonstrated, especially on precipitation forecasts (see e.g. de Haan (2008), Brenot et al. (2013), Bennitt and Jupp (2012), Yan et al. (2009)). However, while mostly precipitation-related studies represent the current focus of research (see Guerova et al. (2016) for a comprehensive summary), other meteorological phenomena can also be investigated by means of GNSS. The number of studies on other meteorological processes is relatively small, covering phenomena such as thunderstorm activity (de Haan (2013)) or fog formation (Stoycheva and Guerova (2015), Aichinger-Rosenberger (2018)). Stoev and Guerova (2018) represent the only investigation on foehn winds using GNSS products to our knowledge, in an initial study for Bulgaria based on observations of Integrated Water Vapour (IWV). The present study, however, develops an entirely new methodology for tackling this problem, by using a novel, data-driven approach.

Foehn winds are a characteristic weather phenomenon in mountainous regions all over the world, especially in the vicinity of prominent mountain ranges like the Alps (where it is typically referred to as Alpine foehn). In general, foehn can be characterized as 'a wind (which is) warmed and dried by descent, in general on the lee side of a mountain' (WMO, 1992). This definition already includes its major characteristics observed in affected areas: strong gusty winds, increasing temperatures and decreasing humidity. While there are many other effects of foehn winds (from social to economic impacts), large wind speeds and gusts are the most critical features from the perspective of operational forecasting and warning systems. In typical foehn valleys like the Reuss Valley (Switzerland) or the Wipp Valley (Austria) wind speeds up to 100 km/h are common, and even up to 200 km/h gusts can be observed at high altitude stations.

Foehn research denotes one of the major topics of (alpine) mountain meteorology (Steinacker, 2006). Despite the fact that the underlying physical processes of foehn have been studied for over a century, still some gaps in knowledge, especially concerning small-scale features, exist. As the classical thermodynamic foehn theory is not able to sufficiently explain all observed foehn events (especially those lacking precipitation), a number of different theories and extensions have been proposed. Furthermore, large observation campaigns like the Mesoscale Alpine Program (MAP) have been conducted and combined with NWP results in order to assess small-scale effects (Gohm and Mayr (2004), Lothon et al. (2006), Drobinski et al. (2007), Mayr et al. (2007)).

Despite these substantial efforts in research, classification and forecasting of foehn both are still challenging tasks. Classification by human expertise still provides the most accurate results, as dedicated experiments, comparing subjective and objective methods, reveal (Mayr et al., 2018). The ability of NWP models to predict foehn is limited by the fact that small-scale features still can not be modelled with sufficient accuracy due to coarse representation of real-world topography (Wilhelm, 2012).

Machine learning (ML) techniques have been a major research topic in atmospheric sciences over the last decade. ML-related approaches of post-processing NWP output, known as model output statistics (MOS), have been shown to significantly enhance operational weather forecasts, see e.g. Glahn and Lowry (1972), Wilks and Hamill (2007) or Hess (2020). ML has also been used to assign uncertainty estimates to forecasts based on deep learning methods applied to previous forecasts (Scher and Messori, 2018). Furthermore, the classification and detection of different weather types has been advanced and automated for different kinds of weather phenomena such as thunderstorms (Perler and Marchand (2009), Manzato (2005)), temperature

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022


Atmospheric Measurement Techniques Discussions

forecasts (Yalavarthi and Shashi (2009)), wind systems (Kretzschmar et al. (2004), Otero and Araneo (2021)) or large-scale weather regimes in general (Deloncle et al., 2007). Common ML methods for such classification problems investigated in former studies are e.g.:

- Random forests: Deloncle et al. (2007)
- Adaptive boosting (AdaBoost): Perler and Marchand (2009), Sprenger et al. (2017)
- Support vector machines : Yalavarthi and Shashi (2009)
- Neural networks: Manzato (2005), Kretzschmar et al. (2004), Otero and Araneo (2021)

Only few authors have used ML approaches for detection and prediction of Alpine foehn yet. Initial studies were carried out by Sprenger et al. (2017), who applied the AdaBoost algorithm to a data set combining weather station observations with NWP output fields from COSMO (Consortium for Small-scale Modeling). They found good performance of the algorithm, obtaining high values for probability of detecting foehn events (88%) and ratio for correct alarms of the algorithm (66%). The most recent study by Mony et al. (2021) showed the feasibility of using ERA5 reanalysis and climate model output instead of NWP output in a similar way as Sprenger et al. (2017).
Contrary to atmospheric science, ML-based approaches have not yet been used as extensively in geodesy. Only very recently, more and more studies involve ML as a tool for problems like discontinuity detection or trend estimation in geodetic time series as well as the prediction of geodetic parameters. Crocetti et al. (2021) showed the feasibility of using ML algorithms for the detection of earthquakes based on GNSS station coordinate time series from Japan. Ruttner et al. (2022) were able to connect raw meteorological parameters with observed GNSS height residuals using a Temporal Convolutional Network (TCN).

The presented study investigates the usability of GNSS tropospheric parameter time series in combination with ML-based classification algorithms for the detection and prediction of foehn events in Switzerland. For this purpose, we utilize a data set spanning eleven years (2010-2020), derived from GNSS observations at sites all over Switzerland as well as a long-term record of foehn observations at the meteorological station Altdorf located in Central Switzerland. The performance of different ML algorithms for foehn prediction and classification is assessed. In a further step, we extend the approach for the application of near real-time (NRT) GNSS products to explore their capabilities to predict foehn events and evaluate the potential of the proposed method for operational weather forecasting.

2 GNSS Meteorology

As already outlined in the introduction, the concept of GNSS Meteorology is based on the fact that electromagnetic signals are delayed by the presence of the Earth's atmosphere. The signal delay is directly proportional to the refractive index n of the atmosphere. In the neutral atmosphere, the refractive index or refractivity N depends on the pressure P (hPa), temperature T

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022


Atmospheric Measurement Techniques Discussions

(K), as well as the water vapor partial pressure e (hPa) (Rüeger, 2002):

$$N = (n-1) \times 10^6 = \frac{77.6890 \cdot P}{T} + \frac{6.3938 \cdot e}{T} + \frac{3.75463 \times 10^5 \cdot e}{T^2}, \quad (1)$$

The term can be split into a hydrostatic (addressing the first term) and a wet portion (addressing the last two terms of Equation 1). The total tropospheric delay experienced by a GNSS signal observed at an elevation e and azimuth direction a is referred to as the Slant Total Delay (STD)

$$\text{STD}(a,e) = \text{ZHD} \cdot mf_h(e) + \text{ZWD} \cdot mf_w(e) + mf_g(e) \cdot [\text{GN} \cdot cos(a) + \text{GE} \cdot sin(a)], \quad (2)$$

where ZHD (Zenith Hydrostatic Delay) represents hydrostatic part, and ZWD the wet part of the signal delay in the zenith direction. In addition, horizontal gradients GN (north-south direction) and GE (east-west direction), accounting for the asymmetry of the atmospheric layers passed by the signal, can be estimated in GNSS processing. In order to map the delays and gradients estimated for the zenith direction to the correct elevation, mapping functions for both parts of the delay ($mf_h(e), mf_w(e)$) and the gradients ($mf_g(e)$) are used.

The total delay in the zenith direction, i.e. the Zenith Total Dealy (ZTD), is the sum of the hydrostatic and wet part

$$\text{ZTD} = \text{ZHD} + \text{ZWD}. \quad (3)$$

ZHD accounts for the major part of the total delay and is largely determined by the atmospheric pressure. It can be modelled with sufficient accuracy from surface pressure observations using, e.g., the formula of Saastamoinen (Saastamoinen, 1972):

$$\text{ZHD} = \frac{0.0022767 \cdot p_s}{1 - 0.00266 \cdot \cos(2\theta) - 0.00028 \cdot H} \quad (4)$$

where p_s is the surface pressure, θ the station latitude, and H is the station height above the geoid.

ZWD represents the main signal of interest for meteorological purposes, as it is directly related to the water vapor content in the air column above the GNSS receiver, and therefore to IWV, via

$$IWV = \kappa(T_m) \cdot ZWD, \quad (5)$$

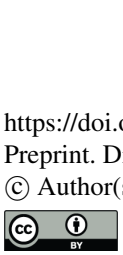
where κ denotes a semi-empirical function depending on the integrated mean temperature T_m. Thus, it shows the same high temporal and spatial variability as water vapor, making precise modelling from meteorological surface observations practically impossible. As a consequence, ZWD is commonly estimated as an unknown in GNSS parameter estimation alongside of station coordinates and the receiver clock error.

3 Data

3.1 GNSS tropospheric products

All investigations presented in this study are based on GNSS troposphere products from the Automated GNSS Network Switzerland (AGNES). The AGNES network was established in 2001 and is maintained by the Swiss Federal Office of To-

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022



Atmospheric Measurement Techniques Discussions

pography (swisstopo) (Brockmann et al., 2002). Currently it consists of 31 GNSS stations, which are visualized in Figure 1. The capabilities of the network were extended to multi-GNSS in 2015 (Brockmann, 2016). Reprocessed, long-term time series

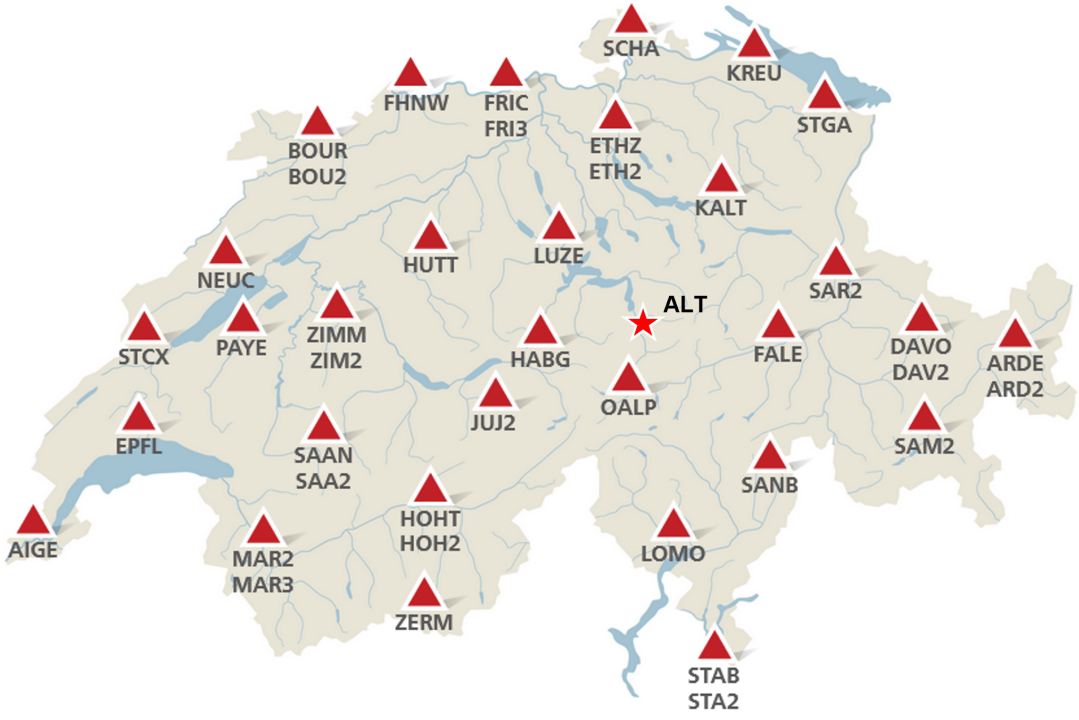


Figure 1. Overview of the AGNES GNSS station network, the meteorological station at Altdorf (ALT) is marked with a star. Picture modified, original taken from: http://pnac.swisstopo.admin.ch/pages/en/agnes.html

of hourly tropospheric delays and gradients are available for the period 1999-2020. A description of the data set as well as details on the reprocessing of GNSS data can be found in e.g. Brockmann (2015). Parts of this reprocessing were carried out in the framework of the second EUREF (International Association of Geodesy Reference Frame Sub-Commission for Europe) Permanent Network (EPN) reprocessing campaign in 2014, where GNSS data from a large number of European stations were reprocessed (Pacione et al., 2017).

We use the later part of the data set (2010-2020) and split it into training (2010-2018) and test data (2019-2020) as input for the proposed prediction/classification algorithms.

3.2 Foehn observations at Altdorf

In order to train a specific ML algorithm and evaluate its performance, a reference data set of foehn observations is needed as the target variable. This study uses time series of 10-minute estimates of foehn index (FI) calculated at the station Altdorf, following

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022


Atmospheric Measurement Techniques Discussions

the approach presented by Dürr (2008). The FI data is provided by the National Meteorological Ground-level Monitoring Network (SwissMetNet, SMN) operated by MeteoSwiss. Currently, data from about ten sites, frequently experiencing foehn winds, are available on an operational level. The index is designed for operational nowcasting and relies on typical foehn predictors such as wind speed and direction, pressure and temperature gradients, and humidity observations at the respective measurement site and surrounding stations. It can return three different integer values: 0 (no foehn), 1 (foehn-mixed air) and 2 (foehn), which are distinguished based on the predictors mentioned above. In an extensive validation against classifications by human experts, the index showed good performance for indices re-calculated back to 1981 (Gutermann et al., 2012). For a detailed description of the calculation algorithm we refer to Dürr (2008) and Gutermann et al. (2012). As we aim for a binary classification (no foehn or foehn), the cases of FI = 1 are treated as nonfoehn events and therefore mapped to value 0. Furthermore, we map the cases of foehn (FI = 2) to the value 1 for the sake of simplicity in all results shown in the following. Then, each hour in the whole data set where at least one 10-minute value indicates foehn is treated as an hour of foehn appearance, and thus a foehn event.

4 Methodology

4.1 Machine Learning Algorithms

In the course of this study, several different ML algorithms are tested in order to investigate their usability for this specific problem and to compare their performance relative to each other. The following algorithms are tested:

- Adaptive Boosting (ADB) (Freund and Schapire, 1997)
- Gradient Boosting (GB) (Friedman, 2001)
- Multilayer Perceptron (MLP) (LeCun et al., 2012)
- Random Forest Classifier (RF) (Breiman, 2001)
- Support Vector Classifier (SVC) (Platt, 1999)
- K-Nearest Neighbor (KNN) (Cover and Hart, 1967)
- Decision Tree (DT) (Breiman et al., 2017)

As a detailed discussion of these algorithms would go beyond the scope of this study, we refer to Hsieh (2009) for a comprehensive overview.

4.2 Feature selection from GNSS time series

The selection of features from GNSS troposphere time series is based on previous investigations on visual detection from time series of different parameters as well as on obvious choices which are expected to be impacted most by foehn conditions.

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022

Atmospheric Measurement Techniques Discussions

Two obvious choices are visualized for December 2019 in Figure 2, namely ZWD (12-hour moving-average) at stations north (KALT, shown in blue) and south of the Alpine ridge (LOMO, shown in orange) and its difference (bottom section, shown in black). In addition, foehn events at Altdorf are shown as color-coded periods (orange). Strong correlation between the contrary trends in ZWD at the two stations and the onset of foehn in Altdorf can be observed. Furthermore, the difference in ZWD between the stations reaches a (negative) maximum in the two extended foehn periods observed ($\sim$ 15.-18. and 19.-21.12.2019). These time series give a first impression how (and from which parameters) foehn events can be detected using GNSS data sets. As this becomes a very demanding task for longer periods (both visually and analytically), ML-techniques are a promising tool to extend and automate such a detection process, with the additional benefit of possibly providing the ability to also predict upcoming events.

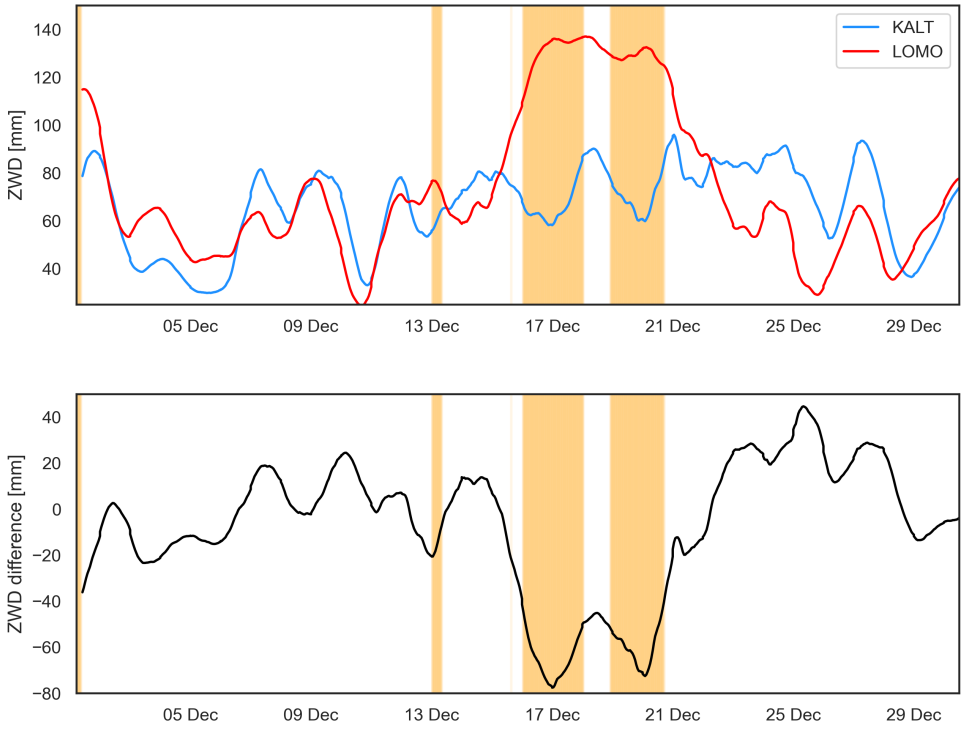


Figure 2. Time series of promising foehn predictors for December 2019. Top: ZWD (12-hour moving-average) from stations KALT (north of Alpine ridge, blue) and LOMO (south of Alpine ridge, orange), Bottom: ZWD difference between KALT and LOMO (black). Observed foehn events at Altdorf are visualized as orange areas.

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022


Atmospheric Measurement Techniques Discussions

4.3 Data preparation

One of the main challenges for ML-based classification algorithms are imbalanced data sets. This imbalance is also strongly present in data sets of foehn observations, since foehn is a rather rare meteorological phenomenon. For the utilized FI data set, the average foehn probability over the 11-year period (2010-2020) amounts to only $\sim 4\%$. Thus, the ratio of under-representation of the minority class (foehn event) compared to the majority class (no foehn event) is as large as 1:25.

4.3.1 Oversampling

A common approach to overcome problems originating from high imbalance in a data set is to oversample the minority class for the training data set. One possible approach to achieve this is the Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002), which we use in this study. The technique creates new (synthetic) instances of the minority class within the training data. For this study, an oversampling of observed foehn hours in the training data set by 25% was conducted using SMOTE, which improves the performance of the applied algorithms by about 20%. The value of 25% oversampling was chosen to achieve a reasonable balance between the advantage of having more usable training events (larger percentage of oversampling) and the fact that foehn is still a rather rare phenomenon (therefore also rare in possible test data sets). All results shown in the following sections are based on pre-processing using this approach.

4.3.2 Shifting of FI time series

In order to assess the suitability of the GNSS troposphere products for operational prediction, a time shift of one hour is applied to the target vector (i.e. FI time series at Altdorf). As operational usage is considered a main goal of the proposed method, the shift is applied for all test cases investigated in this study (also those using post-processed GNSS products). Therefore, each prediction of a foehn event is based on GNSS observations collected one hour before a possible onset of foehn at Altdorf.

4.4 Performance metrics

As already outlined in the last section, the imbalance in data sets of foehn observations is a major obstacle for the application of ML algorithms and the assessment of their performance. For highly imbalanced data, performance metrics typically used in ML studies might not be representative and therefore other options have to be explored. In the case of the present data set, a typical performance measure such as precision alone would not be suitable as it simply compares detected/predicted foehn hours with the observed data for all time steps. Thus, it might happen that an algorithm with optimal precision predicts (almost) no foehn events at all, since this will still result in an optimal performance with regards to precision. In order to overcome these issues we adapt the following performance metrics, as introduced by Barnes et al. (2007) and used in Sprenger et al. (2017). These can be formulated as conditional probabilities P(|) and separated into detection-based:

- Probability of Detection (POD) = P(predicted | observed)
- Probability of False Detection (POFD) = P(predicted | not observed)

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022


Atmospheric Measurement Techniques Discussions

– Missing Rate (MR) = P(not predicted | observed)

and alarm-based metrics:

– Correct Alarm Ratio (CAR) = P(observed | predicted)

– False Alarm Ratio (FAR) = P(not observed | predicted)

– Missing Alarm Rate (MAR) = P(observed | not predicted).

As already visible from the formulations above, POD and CAR are directly connected to each other via the Bayes Theorem. This also implies that there is always a trade-off between those two parameters and therefore only one of them can be optimized, while decreasing the respective other one. Which metric should be optimized strongly depends on the actual application, as already outlined by Sprenger et al. (2017) who argued that alarm-based measures might be more relevant from a forecasters perspective.

In addition, we adopt two measures which represent a kind of mean performance in terms of both CAR and POD for describing the results presented in the next sections. The first one is just the simple average of those two parameters combined, therefore referred to as COMB in the following:

$$\text{COMB} = \frac{\text{POD} + \text{CAR}}{2}. \tag{6}$$

The second adopted metric is based on the F-beta score F_β (Baeza-Yates and Ribeiro-Neto, 1999), which can be formulated using the confusion matrix. The matrix reports the number of false negatives (FNs), false positives (FPs), true negatives (TNs), and true positives (TPs) and thus allows for the calculation of common performance measures in ML such as precision, recall and F-beta score:

$$\text{precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{7}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{8}$$

$$F_\beta = (1+\beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}} \tag{9}$$

The classical F-beta score (F_1, using β = 1) represents the weighted harmonic mean of precision and recall, with a range between 0 (worst case) and 1 (optimal value). As already discussed above, a precision measure might not be representative for results of this study and therefore we use the F_2 score instead of F_1, putting more emphasis on the recall, i.e. on the detection of all foehn events:

$$F_2 = 5 \cdot \frac{\text{precision} \cdot \text{recall}}{(4 \cdot \text{precision}) + \text{recall}}. \tag{10}$$

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022


Atmospheric Measurement Techniques Discussions

Table 1. Default feature/station setup for cross-validation and case studies.

Feature	Used stations
ZWD	Full AGNES network (31 stations)
GN	Full AGNES network (31 stations)
GE	Full AGNES network (31 stations)
ZWD difference	Full AGNES network (all combinations)
ZHD difference	KALT-LOMO, LUZE-STA2, BOU2-LOMO, BOU2-STA2
DOY	—

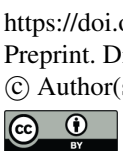
5 Results: Algorithm selection and tuning

This section gives an overview on the process of algorithm selection using a cross-validation approach applied for the whole training data set. Details on this approach are given in Section 5.1. Furthermore the default feature setup (i.e. used GNSS troposphere products) utilized for the algorithm selection as well as for the first case study is outlined in Table 1. It includes ZWD (absolute values and all possible differences between stations) and gradient products (GN and GE) from all available AGNES stations as well as a selection of four ZHD differences which are representative differences between north-south stations in the network. Tests have also been conducted using all possible differences in ZHD (as for ZWD), but no improvement was found using this setup. This might be explained by the fact that ZHD is largely depending on pressure, which typically does not show such small-scale variations as water vapor (and thus ZWD). Therefore, a small number of ZHD differences (i.e. pressure differences) across the Alpine ridge might be sufficient for the ML-based prediction. For the second case study, which uses NRT GNSS products for the prediction, only tropospheric delay products (ZWD, ZHD) are used since gradients products are not available for this data set.

5.1 Choosing optimal ML algorithms

Before carrying out case studies using a specific ML algorithm, the most promising one(s) have to be identified from the list of algorithms given in Section 4.1. Therefore, a cross-validation over the training data set (2010 - 2018) was performed and evaluated using the performance metrics introduced in Section 4.4. For the cross-validation, single years of data are iteratively taken out of the training data set, serving as validation data in order to assess the performance of the outlined algorithms. This is repeated until every year serves once as validation data set. The actual implementation is carried out using the Python package scikit-learn (Pedregosa et al., 2011) and the default settings from the algorithm routines are used in this first evaluation.
Results of the cross-validation are visualized in Figures 3 and 4 and summarized in Table 2. Statistical measures show a strong dependency on the actual foehn probability observed in a specific year, as correlations between black lines (foehn probability) and actual measures for different algorithms (colored lines) reveal. Figure 3 and Table 2 indicate the average performance of all tested algorithms over the nine year validation period. The results indicate best performance for the SVC algorithm in terms of

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022



Atmospheric Measurement Techniques Discussions

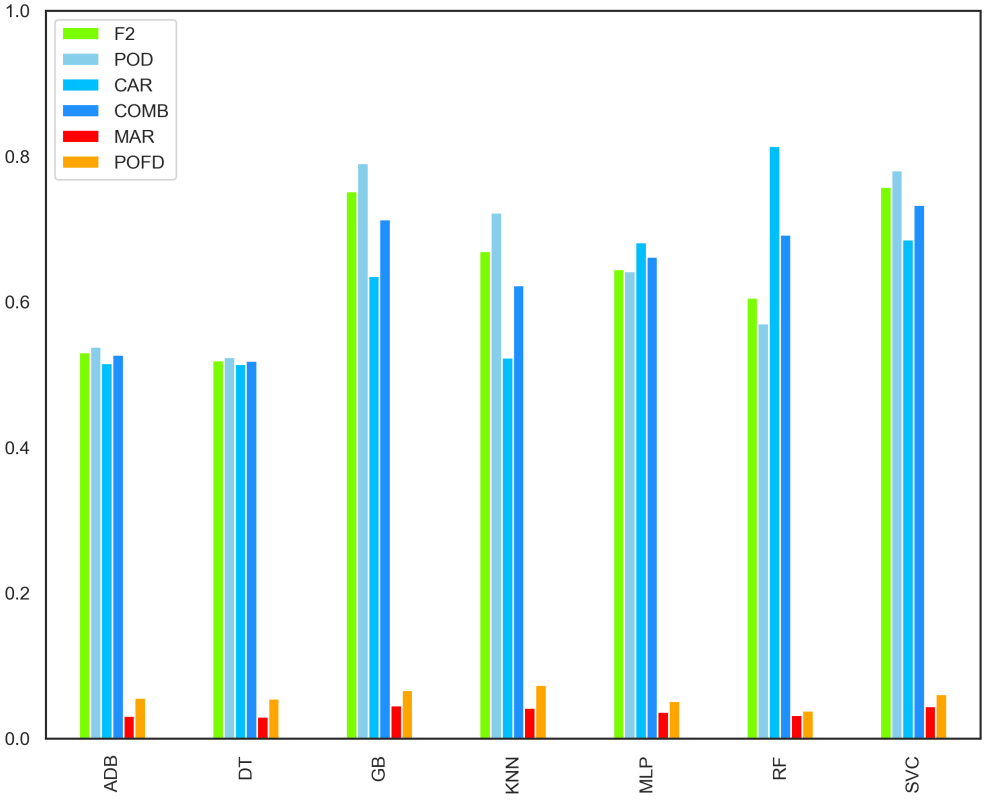


Figure 3. Averaged performance metrics of the cross-validation for all tested algorithms over the nine year training period 2010-2018.

combined measures (COMB and F_2 score). For detection-based measures (POD), the GB algorithm achieves the highest value on average as well as for most single years. The same holds for the RF algorithm in terms of alarm-based measures (CAR), but its detection-based performance is significantly degraded compared to e.g. GB and SVC. Thus we decided to use the GB and SVC algorithms for evaluation in the test case studies, as those are the only ones provided average combined measures of over 70% (see Table 2).

5.2 Hyperparameter tuning

Based on the results of the cross-validation, the GB and SVC algorithm are chosen for in-depth tuning of their hyperparameters. Therefore, a grid-search procedure is conducted, which is an exhaustive search over a subset of manually selected values. The performance of all hyperparameter value combinations is evaluated based on a three-fold cross-validation. Therefore, the train-

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022

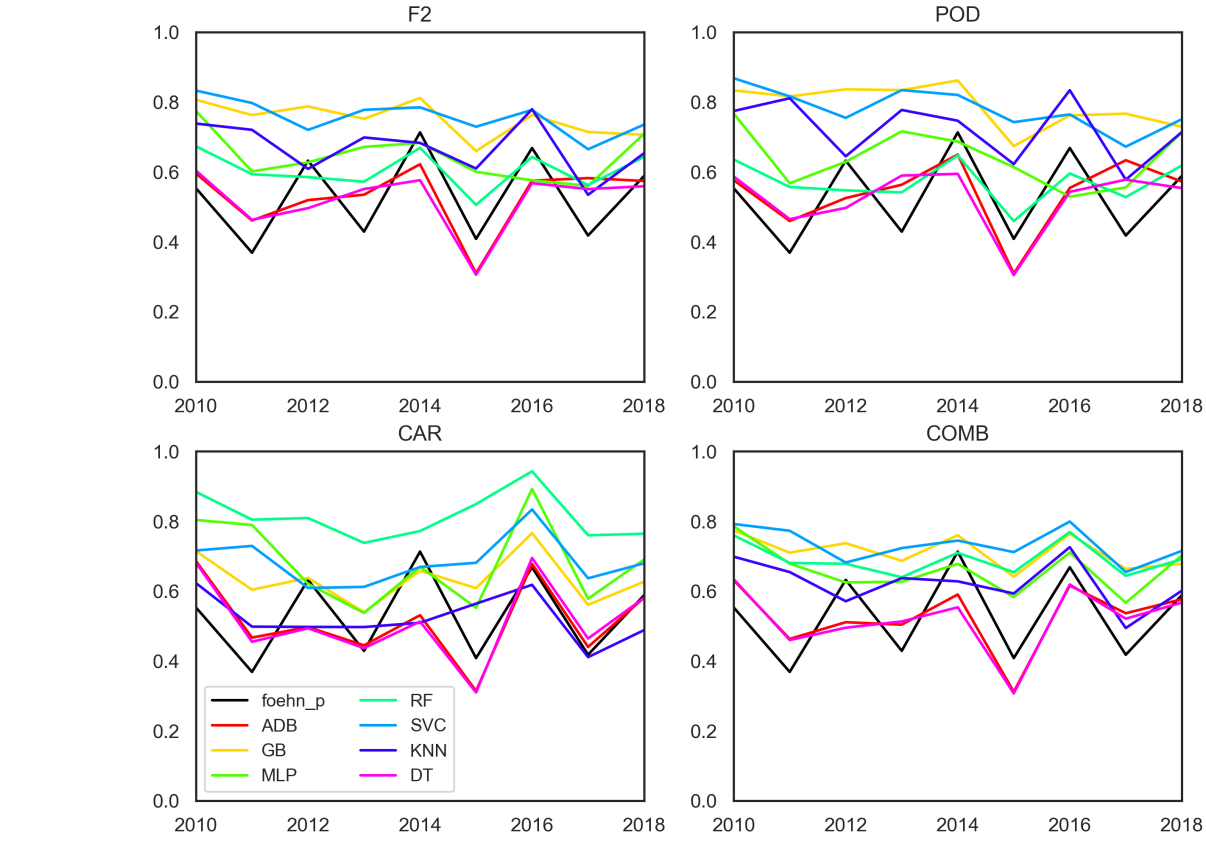


Figure 4. Yearly evolution of performance metrics F_2, POD, CAR and COMB for all tested algorithms. In addition, the black line represents the yearly foehn probability (foehn_p, multiplied by 10 for plotting purposes) calculated from the observed foehn hours at Altdorf.

ing data set (2010-2018) is randomly divided into three folds, where two thirds are used for training while the last third serves for validation. This procedure is repeated three times until each third is used once for validation. All tested hyperparameter values, as well as the best performing value combinations, are summarized in Table 3.

6 Results: Case studies for test period 2019-2020

As the major performance test of the proposed method, two case studies are performed for a dedicated test period, covering the years of 2019 and 2020. Within these case studies, different setups regarding utilized GNSS stations and tropospheric

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022

Atmospheric Measurement Techniques Discussions

Table 2. Averaged performance metrics of the cross-validation for all tested algorithms over the nine year training period 2010-2018.

Algorithm	POD	CAR	COMB	F_2	POFD	MAR
ADB	0.538	0.516	0.527	0.530	0.031	0.056
GB	**0.790**	0.635	0.713	0.751	0.045	0.066
MLP	0.642	0.682	0.662	0.645	0.036	0.051
RF	0.570	**0.814**	0.692	0.605	0.032	**0.038**
SVC	0.780	0.685	**0.733**	**0.758**	0.044	0.061
KNN	0.722	0.523	0.623	0.669	0.042	0.074
DT	0.524	0.514	0.519	0.519	**0.030**	0.054

Table 3. Tuned hyperparameters of both algorithms with their best, tested and default values. For GB, n_estimators stands for the number of boosting stages to perform and max_depth is the maximum depth of the individual regression estimators that limit the number of nodes in the tree. The contribution of each tree is limited by the learning_rate. For SVC, C is the regularization parameter and gamma represents the kernel coefficient.

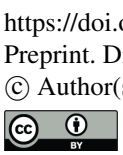
Algorithm	Hyperparameter	Best Value	Tested Values	Default Value
GB	n_estimators	300	[100, 300, 500]	100
	max_depth	5	[3, 5, 8]	3
	learning_rate	0.1	[0.05, 0.1, 0.2]	0.1
SVC	C	10	[0.1, 1, 10, 100, 1000]	1
	gamma	'scale'	[1, 0.1, 0.01, 0.001, 0.0001,'scale']	'scale'

parameters in the feature matrix are investigated. In a final step, we assess the ability of the proposed methods to predict foehn events using near real-time (NRT) troposphere products from previous hours. For all investigations/results shown in the following, the chosen GB and SVC algorithms are trained using nine years (2010 -2018) of hourly troposphere products and evaluated for the test period 2019-2020. Specific settings concerning station and feature setup as well as type of used GNSS products are introduced in the respective sections.

6.1 Case study 1: Reprocessed products

Representing the main results of this study, the first case study includes all available stations and most parameters available from GNSS processing. The exact setup is following the settings obtained from the cross-validation, given in Table 1. Results of the predictions of the test data from both algorithms are visualized in Figure 5 and the respective performance statistics are given in Table 4. The results indicate good performance of the tested algorithms, both providing POD/F_2 values over/at 80% and low POFD/MAR values (3%). The GB prediction achieves a slightly higher CAR (71%) than SVC (67%) and therefore

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022



Atmospheric Measurement Techniques Discussions

Figure 5. Foehn predictions for GB (top) and SVC (bottom) algorithms for the full network run and the test period 2019-2020. Shown are predicted foehn events (orange, yellow), observed FI (red) as well as matches between prediction and observation (green).

also higher COMB values (77% compared to 75%). In addition, Figures 6 and 7 visualize the results of the prediction split up for every month of the two-year data set. Analysis of these monthly results shows most missed alarms (FN, red) in the strongest foehn months, February to April. Extended periods of false alarms (FP, blue) can be seen in December 2020 as well as June 2019. Overall, these findings hold for results from both tested algorithms, although SVC typically produces more false alarms (higher POD but therefore also higher POFD/lower CAR) than GB.

As the GB algorithm gives the opportunity to assess the importance of the used features for the prediction result, we additionally show the 30 most important features in Figure 8. By far the best predictor is the ZWD difference between the stations HABG and SANB, which might outline the importance of high-altitude stations (such as SANB, 1702 m altitude) at the alpine crest.

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022

Atmospheric Measurement Techniques Discussions

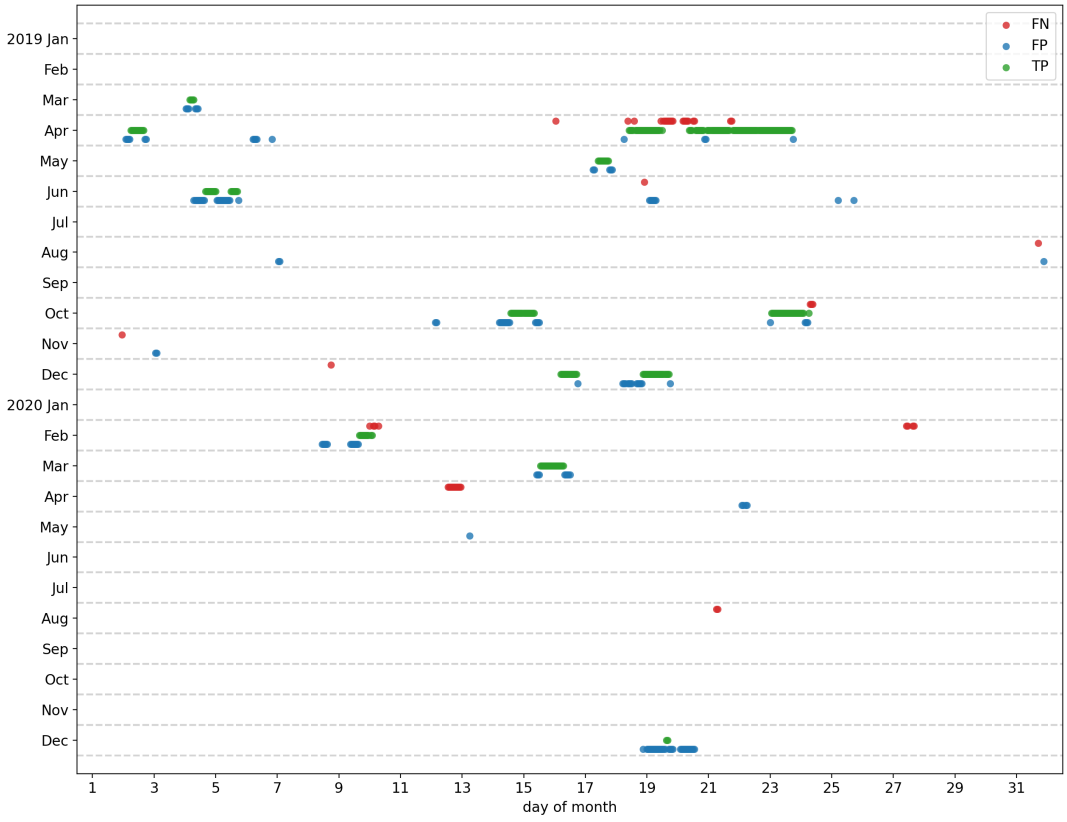


Figure 6. Events of TP (green), FP (blue) and FN (red) for GB predictions, split into individual months of the two-year test data set.

Interestingly, also features from stations far away from Altdorf have significant impact, most prominently ZWD and also gradient products (even for east-west direction) in the Valais area (HOHT and HOH2 stations). This is reasonable due to the fact that typical wind trajectory in the Rhone valley is east-west oriented.

6.1.1 Thinning of feature space

As visible from Table 1, the dimension of the feature space (i.e. the number of columns of the feature matrix) becomes large when the full AGNES network is utilized for the prediction (at least for ZWD and gradient products), which significantly increases the computational effort for algorithm training. In this section, we thus investigate if comparable results can still be obtained when significantly reducing the number of features. Therefore we thinned out the feature space to the top 30 predictors

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022

Atmospheric Measurement Techniques Discussions

Figure 7. Events of TP (green), FP (blue) and FN (red) for SVC prediction, split for all months of the two-year test data set.

Table 4. Performance metrics for the proposed models using post-processed troposphere products and the full feature setup (as shown in Table 1).

Algorithm	POD	CAR	COMB	F_2	POFD	MAR
GB	0.83	0.71	0.77	0.81	0.03	0.03
SVC	0.83	0.67	0.75	0.80	0.04	0.03

290 from the results shown in the prior section (Figure 8). Results obtained from this reduced data set are given in Table 5. While the POD remains unchanged or even increases, the reduction of features leads to a drop of the CAR values by $\sim$ 10% for

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022

Atmospheric Measurement Techniques Discussions

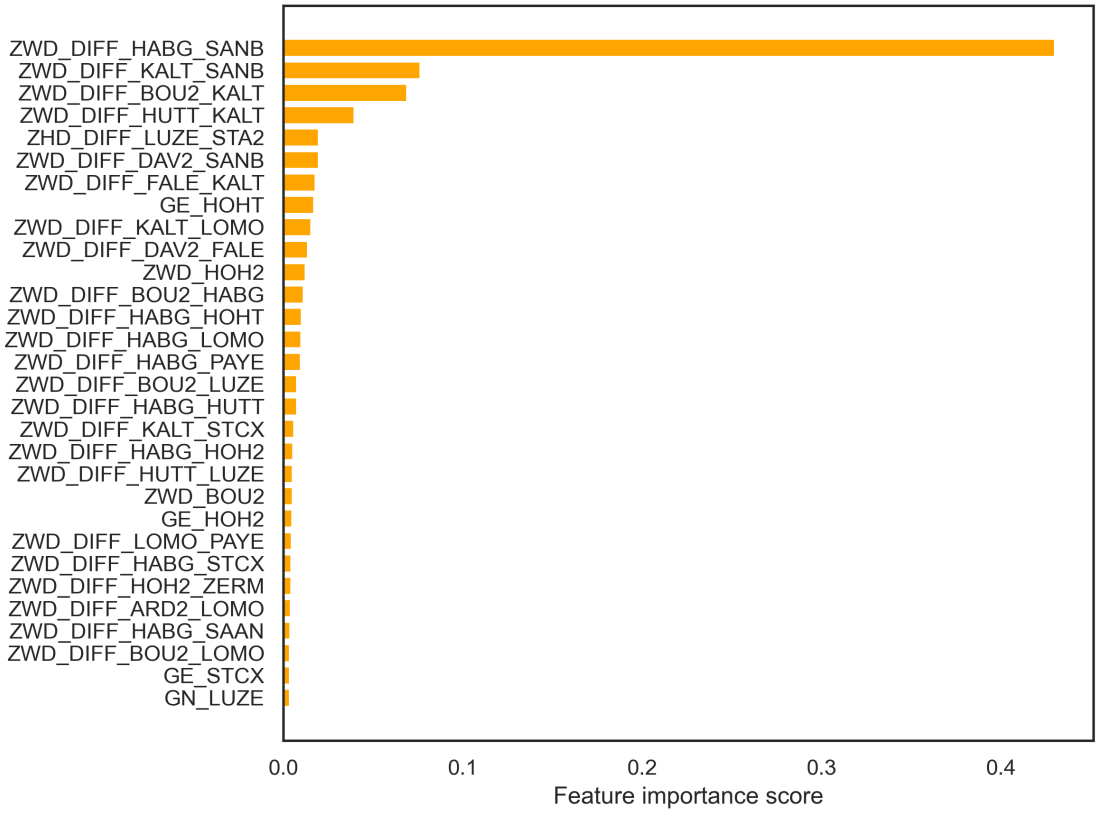


Figure 8. Feature importance score of the 30 strongest predictors for the GB algorithm, when using the full feature setup from post-processing troposphere products.

Table 5. Performance metrics for the proposed models using post-processed troposphere products but only the top 30 features (as shown in Figure 8 from the original setup Table 1).

Algorithm	POD	CAR	COMB	F_2	POFD	MAR
GB	0.83	0.63	0.73	0.78	0.04	0.03
SVC	0.85	0.57	0.71	0.78	0.04	0.03

both algorithms. Therefore, it can be concluded that from a forecasting/prediction perspective, it is not desirable to reduce the number of features (at least not by a large amount as done here).

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022

Atmospheric Measurement Techniques Discussions

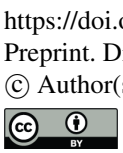
6.2 Case study 2: Prediction using NRT troposphere products

The last major question behind this study is to what degree the proposed method can be used or incorporated into operational forecasting of foehn events. Therefore, the investigations presented in the prior sections are extended by using NRT troposphere products, in order to investigate the suitability of the proposed ML algorithms for operational prediction. These data are currently available in form of NRT tropospheric delays (ZHD, ZWD and ZTD), which are typically provided with a latency of approximately 30-40 minutes after the full hour. Unfortunately, no atmospheric gradients are currently delivered in NRT mode, but an extension is possible and aimed for in the near future. The missing gradient information makes it necessary to train the GB and SVC algorithms again for the dedicated period (2010-2018), but this time only using features related to tropospheric delays (ZWD, ZWD differences, ZHD differences). Results of the prediction using NRT products can be found in Figures 9 (for GB) and 10 (for SVC) as well as in Table 6. The comparison of ML-based predictions with the FI time series shows a

Figure 9. GB predictions for default (GB) and adjusted (GB_adj) run using NRT products. Shown are predicted foehn events (orange, yellow), observed FI (red) and matches (to adjusted prediction, green).

significant decrease in prediction accuracy, especially in terms of CAR for both algorithms. This indicates the importance of gradient parameters for the proposed method, despite the fact that the feature importances are dominated by ZWD differences (see Figure 8). Furthermore, lower quality of ZWD estimates must be taken into consideration as well for the NRT solution, since lower-quality orbit and clock products have to be used for GNSS processing. Nevertheless, combined measures (F_2 and

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022



Atmospheric Measurement Techniques Discussions

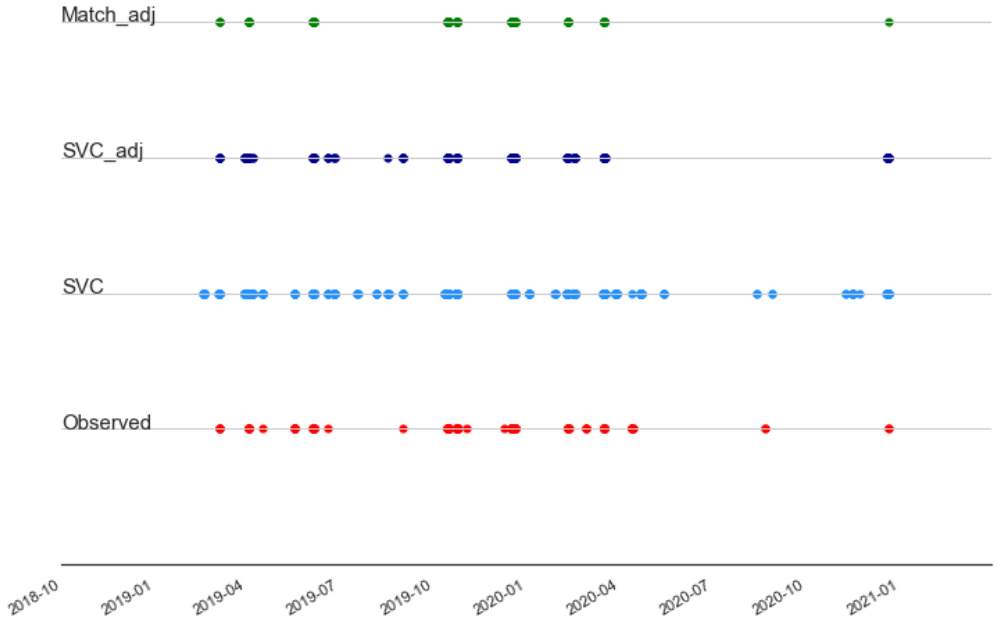


Figure 10. SVC predictions for default (SVC) and adjusted (SVC_adj) run using NRT products. Shown are predicted foehn events (blue, darkblue), observed FI (red) and matches (to adjusted prediction, green).

Table 6. Performance metrics for the proposed models and usage of NRT data from the full station network.

Algorithm	POD	CAR	COMB	F_2	POFD	MAR
GB	0.84	0.37	0.60	0.66	0.04	0.02
GB (adjusted)	0.69	0.60	0.67	0.65	0.04	0.02
SVC	0.85	0.34	0.60	0.65	0.05	0.02
SVC (adjusted)	0.70	0.56	0.63	0.66	0.02	0.01

COMB) are still reaching values between 60-66%. If a more equal performance level for both detection- and alarm-based measures is desired, the initial value of the decision threshold (0.5) used by an algorithm for the binary classification can be adopted. In this case, we used a threshold value of 0.75 to gain more equal statistics for both algorithms (results marked with _adj in Figures 9 and 10), while conserving the combined measures. This higher threshold is able to increase the CAR by ∼20%, given the fact that it erases a considerable amount of false alarms.

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022


Atmospheric Measurement Techniques Discussions

7 Conclusions and outlook

In the present study, the feasibility of using GNSS troposphere products in combination with ML-based classification algorithms for the detection and operational prediction of alpine foehn was analyzed. For this purpose, we made use of a long-term (eleven years) data set of GNSS tropospheric parameters derived for stations of the Swiss AGNES GNSS network as well as FI observations at the SMN-station Altdorf. In the course of an extensive cross-validation over the training data set (2010-2018), seven different classification algorithms were tested. The two best-performing algorithms were then trained using the nine-year training data set already utilized for cross-validation earlier. In a first case study, we evaluated results of foehn classifications/predictions from those two algorithms (GB and SVC) over a two-year test period (2019-2020) at the SMN-station Altdorf. The second case study investigated the usability of NRT GNSS products for foehn prediction, in order to assess the feasibility of these low-latency ($\sim$ 30-40 minutes) data for operational forecasting.

The following main conclusions can be drawn from the presented results:

- ML-based foehn prediction/classification using GNSS troposphere products works well and achieves good performance in terms of both detection-based (POD = 83%, POFD = 4%) and alarm-based (CAR = 67-71%, MAR = 3%) metrics. Results of both algorithms are almost equivalent and comparable to those obtained by Sprenger et al. (2017), who used NWP data as well as observations from the validating measurement site (Altdorf). Thus, the obtained "GNSS-only" results are unexpectedly promising. This fact once more outlines the great potential of GNSS products for meteorology – not only for precipitation-related phenomena – as well as the benefits of using them in ML-based approaches.

- Out of the seven algorithms tested in the nine-year cross-validation, GB and SVC provided the best average performance in terms of performance metrics mentioned before. Therefore they were selected for further hyperparameter tuning and usage in the case studies carried out.

- Most promising results can be obtained if the full station network can be utilized. A reduction of the feature space to the 30 most important features from the full-network run shows similar POD values, but results in a degradation of CAR values by $\sim$10%. Still, a careful thinning in the feature space (much higher number than 30 features remaining) might be possible without losing much critical information. Most important predictors include ZWD differences from GNSS stations nearby Altdorf as well as gradient products from selected stations, which include e.g. stations in the Valais area.

- Similar problems (feature reduction) as outlined above are present when using NRT data for operational prediction, as currently only tropospheric delays are delivered in NRT products. Furthermore, the quality of the prediction results also varies with the quality of the troposphere products, which is significantly lower for NRT data. This is visible from the comparison between post-processed delay-only and NRT runs. Similarly as the for the dedicated feature thinning, main degradation of the results is experienced for the alarm-based measures. Still, the degradation can be mitigated to some extent by a dedicated tuning of the decision threshold in the classification algorithm (increase/decrease of the default 0.5 value). Finally, we except that the NRT prediction will also benefit from gradient products as soon as they become available in an operational manner.

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022


Atmospheric Measurement Techniques Discussions

– Choosing the optimal performance metrics and appropriate pre-processing denotes a key task in ML-based prediction algorithms, especially when working with such a highly imbalanced data set as in this study. The actual choice for the most important metric(s) strongly depends on the actual application of the prediction method, deciding whether detection- or alarm-based measures should be preferred. Within this study we tried to tune the algorithms for an optimal balance between both metric-types and leave a possible decision to the potential users. However, as already outlined before, there exists a trade-off between POD and CAR and therefore an optimization towards one metric will always result in a shortcoming towards the other.

Overall, these initial results are very promising and therefore the developed method might already aid the meteorological community as an additional tool for foehn classification and/or prediction. Nevertheless, a number of enhancements can still be achieved through more detailed investigations in future studies. Possible improvements of the method we aim to investigate in the future would be:

– Enhance the nowcasting capabilities of the proposed method by including gradient products in the NRT prediction.

– Use troposphere products from relevant stations of GNSS networks from neighbouring countries (Italy, Austria, Germany...) and the small-scale COGEAR network in the Valais area.

– Perform spatial interpolation of tropospheric parameters from neighbouring stations to the location of the SMN-station Altdorf (using e.g. collocation methods) and evaluate their benefit for the prediction.

– Performing a more extensive grid-search for hyperparamter tuning of the used algorithms or try new (possibly more sophisticated deep learning) algorithms.

Code availability. Source code is available upon request from the corresponding author.

Author contributions. MAR developed the research idea, software implementation and prepared the manuscript (including formal analysis, visualization, and writing). EB produced and provided the long-term time series of GNSS troposphere products used for this research. LC prepared selected visualizations for the manuscript. All co-authors contributed to the discussion of results as well as reviewing and editing the manuscript.

Competing interests. The authors declare that no competing interests are present.

Acknowledgements. The authors would like to thank swisstopo for providing GNSS troposphere products from the AGNES network and MeteoSwiss for providing foehn index observations for the station Altdorf.

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022

Atmospheric Measurement Techniques Discussions

References

Aichinger-Rosenberger, M.: Usability of high-resolution GNSS-ZTD data in the AROME model, Master's thesis, University of Innsbruck, https://diglib.uibk.ac.at/urn:nbn:at:at-ubi:1-27392, 2018.

Baeza-Yates, R. A. and Ribeiro-Neto, B.: Modern Information Retrieval, Addison-Wesley Longman Publishing Co., Inc., USA, 1999.

Barnes, L. R., Gruntfest, E. C., Hayden, M. H., Schultz, D. M., and Benight, C.: False Alarms and Close Calls: A Conceptual Model of Warning Accuracy, Weather and Forecasting, 22, 1140 – 1147, https://doi.org/10.1175/WAF1031.1, 2007.

Bennitt, G. V. and Jupp, A.: Operational Assimilation of GPS Zenith Total Delay Observations into the Met Office Numerical Weather Prediction Models, Monthly Weather Review, 140, 2706–2719, https://doi.org/10.1175/MWR-D-11-00156.1, 2012.

Bevis, M., Businger, S., Herring, T. A., Anthes, R. A., and Ware, R. H.: GPS Meteorology: Remote Sensing of Atmospheric Water Vapor Using the Global Positioning System, Geophys. Mag., 34, 359–425, 1992.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J.: Classification and regression trees, Routledge, 2017.

Brenot, H., Neméghaire, J., Delobbe, L., Clerbaux, N., De Meutter, P., Deckmyn, A., Delcloo, A., Frappez, L., and Van Roozendael, M.: Preliminary signs of the initiation of deep convection by GNSS, Atmos. Chem. Phys., 13, 5425—-5449, 2013.

Brockmann, E.: Reprocessed GNSS tropospheric products at swisstopo, in: GNSS4SWEC workshop, Thessaloniki, May 11-14, 2015.

Brockmann, E.: National Report of Switzerland, in: EUREF-Symposium, San Sebastian, Spain, May 25-27, 2016.

Brockmann, E., Grünig, S., Schneider, D., Wiget, A., and Wild, U.: National report of Switzerland: Introduction and first applications of a Real-Time Precise Positioning Service using the Swiss Permanent Network AGNES, In Torres, J.A. and H. Hornik (Eds): Subcommission for the European Reference Frame (EUREF). EUREF Publication No. 10. Mitteilungen des Bundesamtes für Kartographie und Geodäsie, 23, 272–276, 2002.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P.: SMOTE: Synthetic Minority Over-sampling Technique, Journal of Artificial Intelligence Research, 16, 321–357, https://doi.org/10.1613/jair.953, 2002.

Cover, T. and Hart, P.: Nearest neighbor pattern classification, IEEE Transactions on Information Theory, 13, 21–27, https://doi.org/10.1109/TIT.1967.1053964, 1967.

Crocetti, L., Schartner, M., and Soja, B.: Discontinuity Detection in GNSS Station Coordinate Time Series Using Machine Learning, Remote Sensing, 13, https://doi.org/10.3390/rs13193906, 2021.

de Haan, S.: Meteorological applications of a surface network of Global Positioning System receivers, Ph.D. thesis, Wageningen University, 2008.

de Haan, S.: Assimilation of GNSS ZTD and radar radial velocity for the benefit of very-short-range regional weather forecasts, Q. J. R. Meteorol. Soc., 139, 2097–2107, 2013.

Deloncle, A., Berk, R., D'Andrea, F., and Ghil, M.: Weather Regime Prediction Using Statistical Learning, Journal of The Atmospheric Sciences - J ATMOS SCI, 64, 1619–1635, https://doi.org/10.1175/JAS3918.1, 2007.

Drobinski, P., Steinacker, R., Richner, H., Baumann-stanzer, K., Beffrey, G., Bénech, B., Berger, H., Chimani, B., Dabas, A., Dorninger, M., Dürr, B., Flamant, C., Frioud, M., Furger, M., Gröhn, I., Gubser, S., Gutermann, T., Häberli, C., Häller-Scharnhost, E., Jaubert, G., Lothon, M., Mitev, V., Pechinger, U., Piringer, M., Ratheiser, M., Ruffieux, D., Seiz, G., Spatzierer, M., Tschannett, S., Vogt, S., Werner, R., and Zängl, G.: Föhn in the Rhine Valley during MAP: A review of its multiscale dynamics in complex valley geometry, Quarterly Journal of the Royal Meteorological Society, 133, 897–916, 2007.

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022


Atmospheric Measurement Techniques Discussions

Dürr, B.: Automatisiertes Verfahren zur Bestimmung von Föhn in Alpentälern, Arbeitsberichte der MeteoSchweiz, 223, 22 pp, 2008.

Freund, Y. and Schapire, R. E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, Journal of Computer and System Sciences, 55, 119–139, https://doi.org/https://doi.org/10.1006/jcss.1997.1504, 1997.

Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, 29, 1189–1232, http://www.jstor.org/stable/2699986, 2001.

Glahn, H. R. and Lowry, D. A.: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting, Journal of Applied Meteorology and Climatology, 11, 1203 – 1211, https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2, 1972.

Gohm, A. and Mayr, G. J.: Hydraulic aspects of föhn winds in an Alpine valley, Quarterly Journal of the Royal Meteorological Society, 130, 449–480, https://doi.org/https://doi.org/10.1256/qj.03.28, 2004.

Guerova, G., Jones, J., Dousa, J., Dick, G., Haan, S., Pottiaux, E., Bock, O., Pacione, R., Elgered, G., Vedel, H., and Bender, M.: Review of the state-of-the-art and future prospects of the ground-based GNSS meteorology in Europe, Atmospheric Measurement Techniques Discussions, 9, https://doi.org/10.5194/amt-9-5385-2016, 2016.

Gutermann, T., Dürr, B., Richner, H., and Bader, S.: Föhnklimatologie Altdorf: die lange Reihe (1864-2008) und ihre Weiterführung, Vergleich mit anderen Stationen, https://doi.org/10.3929/ethz-a-007583529, 2012.

Hess, R.: Statistical postprocessing of ensemble forecasts for severe weather at Deutscher Wetterdienst, Nonlinear Processes in Geophysics, 27, 473–487, https://doi.org/10.5194/npg-27-473-2020, 2020.

Hsieh, W. W.: Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels, Cambridge University Press, https://doi.org/10.1017/CBO9780511627217, 2009.

Kretzschmar, R., Eckert, P., Cattani, D., and Eggimann, F.: Neural Network Classifiers for Local Wind Prediction, Journal of Applied Meteorology - J APPL METEOROL, 43, 727–738, https://doi.org/10.1175/2057.1, 2004.

LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K.-R.: Efficient backprop, in: Neural networks: Tricks of the trade, pp. 9–48, Springer, 2012.

Lothon, M., Druilhet, A., Bénech, B., Campistron, B., Bernard, S., and Said, F.: Experimental study of five föhn events during the Mesoscale Alpine Programme: From synoptic scale to turbulence, Quarterly Journal of the Royal Meteorological Society, 129, 2171 – 2193, https://doi.org/10.1256/qj.02.30, 2006.

Manzato, A.: The Use of Sounding-Derived Indices for a Neural Network Short-Term Thunderstorm Forecast, Weather and Forecasting, 20, 896 – 917, https://doi.org/10.1175/WAF898.1, 2005.

Mayr, G., Plavcan, D., Armi, L., Elvidge, A., Grisogono, B., Horvath, K., Jackson, P., Neururer, A., Seibert, P., Steenburgh, J. W., Stiperski, I., Sturman, A., Željko Večenaj, Vergeiner, J., Vosper, S., and Zängl, G.: The Community Foehn Classification Experiment, Bulletin of the American Meteorological Society, 99, 2229 – 2235, https://doi.org/10.1175/BAMS-D-17-0200.1, 2018.

Mayr, G. J., Armi, L., Gohm, A., Zängl, G., Durran, D. R., Flamant, C., Gabersek, S., Mobbs, S. D., Ross, A. B., and Weissmann, M.: Gap flows: Results from the Mesoscale Alpine Programme, Quarterly Journal of the Royal Meteorological Society, 133, 881–896, 2007.

Mony, C., Jansing, L., and Sprenger, M.: Evaluating Foehn Occurrence in a Changing Climate Based on Reanalysis and Climate Model Data Using Machine Learning, Weather and Forecasting, 36, 2039 – 2055, https://doi.org/10.1175/WAF-D-21-0036.1, 2021.

Otero, F. and Araneo, D.: Zonda wind classification using machine learning algorithms, International Journal of Climatology, 41, E342–E353, https://doi.org/https://doi.org/10.1002/joc.6688, 2021.

Pacione, R., Araszkiewicz, A., Brockmann, E., and Dousa, J.: EPN-Repro2: A reference GNSS tropospheric data set over Europe, Atmospheric Measurement Techniques, 10, 1689–1705, https://doi.org/10.5194/amt-10-1689-2017, 2017.

https://doi.org/10.5194/amt-2022-33
Preprint. Discussion started: 4 February 2022

Atmospheric Measurement Techniques Discussions

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Perler, D. and Marchand, O.: A Study in Weather Model Output Postprocessing: Using the Boosting Method for Thunderstorm Detection, Weather and Forecasting, 24, https://doi.org/10.1175/2008WAF2007047.1, 2009.

Platt, J. C.: Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods, in: Advances In Large Margin Classifiers, pp. 61–74, MIT Press, 1999.

Ruttner, P., Hohensinn, R., D'Aronco, S., Wegner, J. D., and Soja, B.: Modeling of Residual GNSS Station Motions through Meteorological Data in a Machine Learning Approach, Remote Sensing, 14, https://doi.org/10.3390/rs14010017, 2022.

Rüeger, J.: Refractive Index Formulae for Radio Waves, Proc. FIG XXII International Congress, Washington, D. C., 2002.

Saastamoinen, J.: Contributions to the theory of atmospheric refraction., Bull. Geodesique, 105, 13–34, 1972.

Scher, S. and Messori, G.: Predicting weather forecast uncertainty with machine learning, Quarterly Journal of the Royal Meteorological Society, 144, 2830–2841, https://doi.org/https://doi.org/10.1002/qj.3410, 2018.

Sprenger, M., Schemm, S., Oechslin, R., and Jenkner, J.: Nowcasting Foehn Wind Events Using the AdaBoost Machine Learning Algorithm, Weather and Forecasting, 32, 1079 – 1099, https://doi.org/10.1175/WAF-D-16-0208.1, 2017.

Steinacker, R.: Alpiner Foehn - a new verse to an old song, Promet, 32, 3–10, 2006.

Stoev, K. and Guerova, G.: Use of GNSS tropospheric products to study the foehn in Sofia, Talk: EMS Annual Meeting: European Conference for Applied Meteorology and Climatology 2018 | 3–7 September 2018 | Budapest, Hungary, 2018.

Stoycheva, A. and Guerova, G.: Study of fog in Bulgaria by using the GNSS tropospheric products and large scale dynamic analysis, Journal of Atmospheric and Solar-Terrestrial Physics, 133, 87–97, https://doi.org/https://doi.org/10.1016/j.jastp.2015.08.004, 2015.

Wilhelm, M.: COSMO-2 Model Performance in Forecasting Foehn: a Systematic Process-oriented Verification, Veröffentlichungen der MeteoSchweiz, 89, 64pp, 2012.

Wilks, D. S. and Hamill, T. M.: Comparison of Ensemble-MOS Methods Using GFS Reforecasts, Monthly Weather Review, 135, 2379 – 2390, https://doi.org/10.1175/MWR3402.1, 2007.

WMO: International Meteorological Vocabulary. WMO No. 182, 1992.

Yalavarthi, R. and Shashi, M.: Atmospheric Temperature Prediction using Support Vector Machines, International Journal of Computer Theory and Engineering, 1, 55–58, https://doi.org/10.7763/IJCTE.2009.V1.9, 2009.

Yan, X., Ducrocq, V., Jaubert, G., Brousseau, P., Poli, P., Champollion, C., Flamant, C., and Boniface, K.: The benefit of GPS zenith delay assimilation to high-resolution quantitative precipitation forecast. A case-study from COPS IOP 9., Q. J. R. Meteorol. Soc., 135, 1788—-1800, 2009.