Evaluating foehn development in a changing climate using machine learning and investigating the impact of foehn on forest fire occurrence and severity

Master Thesis

Author(s):
Mony, Christoph

Publication date:
2020-12

Permanent link:
https://doi.org/10.3929/ethz-b-000594509

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Evaluating foehn development in a changing climate using machine learning and investigating the impact of foehn on forest fire occurrence and severity

Master's Thesis

Christoph Mony

December 23, 2020

Advisors: Dr. Michael Sprenger[1], Lukas Jansing[1], Prof. Dr. Josef Jochum[2]

[1]Department of Environmental Systems Science, ETH Zurich

[2]Department of Physics, EKU Tuebingen

Acknowledgments

This work concludes my studies of Physics at EKU Tuebingen and ETH Zurich. During my studies, I learned a lot (even though I feel that there is still so much out there) and collected many great experiences. Nevertheless, what fills me with gratitude the most is the amount of support from professors, colleagues, and friends I received along this journey.

Here, in particular, I would like to thank the people that made this final piece of work possible. First, I want to thank Prof. Dr. Heini Wernli and Dr. Michael Sprenger for offering me the chance to work at ETH Zurich during the last months. Second, I want to express my gratitude to Prof. Dr. Josef Jochum for supporting me as my supervisor at EKU Tuebingen. Especially during those uncertain and stressful times of a pandemic, I felt that they all handled the novel challenges of supervision exceptionally well.

Over the last year, I encountered many people who contributed to the final form of this thesis. Certainly, my advisors at ETH Zurich, Dr. Michael Sprenger and Lukas Jansing, had the most considerable impact on my work. I appreciated their critical questions and feedback during meetings, was glad about being pushed to presenting my work on various occasions, and felt honored about being given the freedom to explore various approaches (many of which did not work). Besides, I want to thank Dr. Matthias Roethlisberger, who happily shared his vast knowledge about climate models and statistical testing with me. Also, I would like to mention Dr. Lukas Gudmundsson, who helped me out with his solid understanding of statistical learning theory.

During the second part of the thesis, I will always keep my visit at the WSL institute in Cadenazzo, Ticino, in good memory. Dr. Boris Pezzatti and Dr. Marco Conedera acted as my hosts and provided me with sound expert knowledge about forest fires and potential foehn-fire interaction mechanisms. Additionally, they gave me a first glimpse into the Swiss-Italian culture.

Not being forgotten shall be all of the other group members, which immensely helped me to settle into the new environment of ETH and Switzerland during various lunch and coffee breaks. Being it learning through Heini that not all Swiss people like cheese fondue, getting to know from Raphael what "s'Badi" is, or being asserted by Lukas that I am not the only one having trouble understanding people from the Valais: Sometimes, it is about the little things that stick with us.

This work was financially supported by the PROMOS scholarship of the German Academic Exchange Service for the first half and by the scholarship of the Hein-Moeller-Foundation for the second half of the year. Without their

benevolence, my time in Zurich would not have been possible, and I want to thank both institutions for that.

Finally, I would like to mention some people that backed me beyond my Master's thesis. My parents supported me throughout all of my studies and did not hesitate to support my move to Zurich, which I am deeply grateful for. Further, I want to thank my girlfriend Janine, who acted as my emotional anchor and would force me into a normal rhythm again after my trips of ruining my sleep schedule by working till past midnight. My conversations with my friends Pano and Jonas often allowed me to see things from a more positive perspective, and quarantine would have been only half as endurable without their inexhaustible motivation to ignite others for new projects. Moreover, a big thanks to Jogi and Christian for their radical honesty with me about some of my quirks, giving me the mental slap I sometimes need. Lastly, a big shout-out to Julia, Sandro, Wouter, Frida, Lennart, Izi, and Lovisa for lessons about Switzerland and Swiss culture, all the discussions about the way of living, and our fantastic hiking/cycling/skiing trips.

As obtained from a discussion with Jonas, I thoroughly believe in Isaac Newton's words that our progress is solely possible because we are standing on the shoulders of giants. That is not only true for academia but also our private lives. To all the giants in my life,

Thank You.

Abstract

Foehn winds have a substantial impact on forest fires all over the world. However, two questions have remained unanswered to the foehn community over the last years. First, while accurate forecasts are viable at the timescale of numerical weather prediction, the long-term evolution of monthly foehn occurrence is still open to debate. Second, while many studies cited the importance of foehn winds on forest fires in single case studies, a definite quantitative link between foehn winds and forest fires prevailed missing. Hence, the scientific contribution of this work is divided into two parts.

In the first part, we explored the possibilities of employing machine learning algorithms to predict foehn within Switzerland from its synoptic fingerprint in climate simulations. Here, we used variables from the ERA-Interim reanalysis and the CESM simulation as inputs for our models. We trained on ERA-Interim data to recognize foehn, then verified the results on a CESM simulation of present-day climate, and finally, predicted foehn on a future warming climate CESM simulation. The best generalization between ERAI and CESM was obtained by including the present-day simulation in the training procedure and simultaneously optimizing two objective functions, namely the negative log loss and squared mean loss, on both datasets, respectively. The model verification showed validity of our approach for most of the months. Finally, we found that south foehn in Altdorf is expected to become more common during spring, while north foehn in Lugano is expected to become more common during two summer months.

In the second part, we analyzed forest fires of the past 40 years and linked them with foehn occurrence from a climatological perspective. We found that foehn duration and foehn strength substantially increase both the number and the severity of forest fires. In detail, we observed that if a day showed foehn presence, it was associated with a 3.4-fold increase in numbers of fires in contrast to a day without foehn presence. Furthermore, if a foehn wind set in during the six hours after fire ignition, it increased the median burned area of such fires by a significant factor of three compared to fires without foehn occurrence.

While we developed and tested both methodologies within Switzerland due to the vast availability of foehn and forest fire data, we extensively documented our approaches. Therefore, we encourage other researchers to apply these frameworks also to foehn winds in other regions of the world.

Contents

Chapter 1

Introduction

Forest fires impose a considerable potential threat to humans, animals, and the environment. Although fires are an integral element for many ecosystems and their biodiversity (Robinne et al., 2018), single events might have tremendous consequences for local communities. Often it is argued that fires are associated with occurrences of exceptional weather conditions, e.g., droughts or heatwaves (Littell et al., 2016; Turco et al., 2017; Fink et al., 2004). In order to influence fires, such an event must affect at least one of the following pillars: fuel, oxygen, or temperature (Thomas et al., 2010).

For this reason, in mountainous regions (e.g., the Alps), an additional impact upon forest fires is imposed by another, more local meteorological condition: foehn winds. In short, foehn winds are strong, warm, and dry downslope winds (Richner and Hächler, 2013). Consequently, foehn winds can act upon fires by drying out entire forest areas (Pezzatti et al., 2016) or driving the fire spread itself (Sharples et al., 2010). Many studies and books so far have stressed the impact of foehn on fires in Switzerland (Richner and Hächler, 2013; Zumbrunnen et al., 2009; Sprenger et al., 2016; Pezzatti et al., 2016). However, also in other parts of the world, foehn or foehn-like winds have been shown to affect fires. For example, Schroeder et al. (1964), Keeley (2004), Westerling et al. (2004) and Guzman-Morales et al. (2016) stressed the influence of the Santa Ana winds in California on several severe wildfire events and Sharples et al. (2010) have linked foehn winds in south-eastern Australia to bush fires.

It becomes apparent that an accurate foehn forecast and knowledge about the expected impact of foehn on forest fires are crucial factors for early risk mitigation. For this reason, this work is structured in two parts.

In the first part, we assessed the expected long-term evolution of foehn winds. Currently, it is viable to forecast foehn at the timescale of numerical weather prediction, which is usually limited to several days (Drechsel

and Mayr, 2008). However, in the long-term, it remained unclear which influence the current warming climate will have on the frequency of foehn winds. With this work, we attempted to close this gap for foehn winds in Switzerland. Moreover, we propose a methodological approach that could, in principle, be applied worldwide. For this part, we aim to give an introduction to foehn and its current way of being forecasted under Section 1.1.

In the second part, we investigated the tangible impact of foehn winds on forest fires. The literature has often cited the impact of foehn in single case studies (Richner and Hächler, 2013; Sprenger et al., 2016; Carrega, 1991). However, it remained unresolved how foehn winds influence forest fires from a climatological perspective. For this reason, we quantified how foehn winds affected the number of forest fires and their severity in Switzerland over the last 40 years. Therefore, we introduce the current fire situation in Switzerland and potential foehn-fire interaction mechanisms in Section 1.2.

Lastly, this work follows a modular structure for both parts. In Chapter 1, we introduce the main concepts about foehn winds and fires. In Chapter 2, we explain the conceptual idea behind each part's analysis and present the used algorithms. In Chapter 3, we present the data and necessary preprocessing steps for our analysis. In Chapter 4, we explain the implementation of our algorithms in detail. In Chapter 5, we present our results and discuss them extensively. In Chapter 6, we conclude by summarizing our main findings and giving an outlook to future work.

1.1 Future development of foehn winds

Since humans have started to settle in the Alps, foehn winds have affected their lives. Foehn is a phenomenon that occurs in mountainous regions worldwide, for example, in the European Alps or the Rocky Mountains. Here, the word "foehn" is often referred to as a generic term for a strong, warm, and dry downslope wind (Richner and Hächler, 2013). A more formal definition can be found in WMO (1992) as "wind [which is] warmed and dried by descent, in general on the lee side of a mountain". Depending on the location, foehn winds may be known under a different name like Chinook in North America (Oard, 1993), Nor'wester in New Zealand (Mcgowan and Sturman, 1996; Simpson et al., 2014) or Puelche in South America (Beusch et al., 2018). Even local nicknames like "snow eater" or "grape cooker" have been given to the phenomenon.

Those nicknames originate from the typical characteristics of foehn winds, which are shown for an exemplary case in Figure 1.1. The rapid increase in temperature and wind speed, accompanied by the drop in relative humidity and orientation of the wind direction along the valley axis, characterize a typical foehn event (Richner and Hächler, 2013).

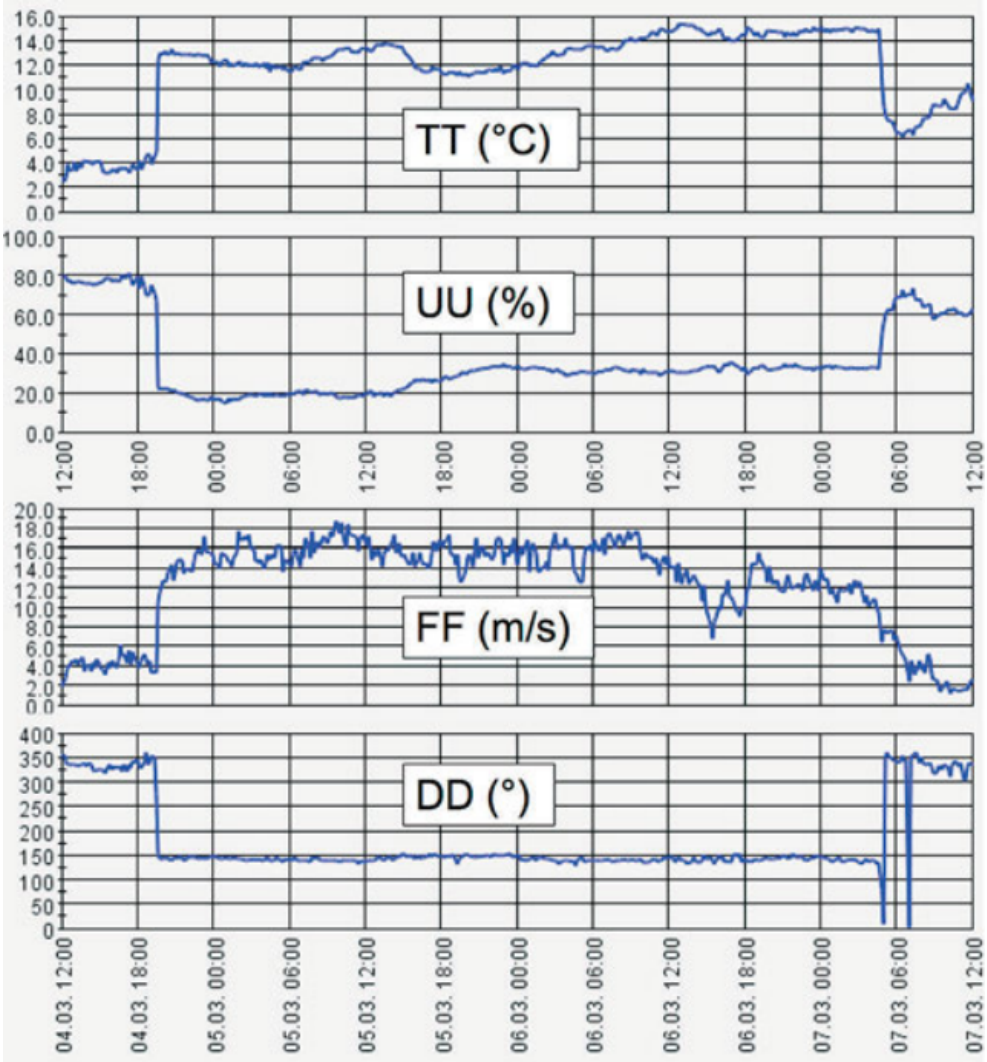


Figure 1.1: Shown are the development of the temperature (TT), relative humidity (UU), wind speed (FF), and wind direction (DD) during a foehn event, which was prevalent in Altdorf from the 04/03/2013 19:00 until 07/03/2013 5:00. Figure taken from Richner et al. (2014).

For these characteristics, foehn is known to not only to have an impact on fires but also on the aviation industry, snowmelt in spring, agriculture, and air pollution (Steinacker, 2006; Richner and Hächler, 2013). Due to its potentially destructive nature, in the next step, we try to develop a better understanding of the mechanisms and causes of foehn.

1.1.1 Mechanisms and causes of foehn

In this section, we first look at the mesoscale mechanisms observed during foehn before moving on to the synoptic situation which causes foehn[1]. Here, we aimed to briefly summarize the last decades' findings while referring to more in-depth literature in the process.

Mesoscale characteristics

At the mesoscale, foehn is divided into two different types, depending on the most prominent warming mechanism of the air (Richner and Hächler, 2013). On the one hand, for "Foehn Type I" (also "Swiss foehn"), wet air rises on the windward side of the mountain through forced convection. Here, the air cools dry-adiabatically due to the ascent. As soon as saturation is reached, condensation and precipitation set in, and the air gains latent heat. After reaching crest height, the air begins to descend on the lee side. Due to the descent, the air begins to heat dry-adiabatically. Over time, the strong and warm flow erodes the colder air, often present in the valley (referred to as "cold pool"). As soon as the cold pool is eroded at a station, the typical foehn characteristics shown in Figure 1.1 emerge (Sprenger et al., 2016). On the other hand, for "Foehn Type II" (also "Austrian foehn"), air in the lower levels of the windward side is blocked orographically. Then, only dry air from the higher levels descends into the lee, heats dry-adiabatically, and replaces the cold pool in the valley. Both types are depicted in Figure 1.2. Most of the time, Austrian foehn is the main warming mechanism behind foehn (Steinacker, 2006; Miltenberger et al., 2016). A more in-depth discussion of both types and more potential warming mechanisms can be found in Richner and Hächler (2013), Ólafsson (2005) and Elvidge and Renfrew (2016).

For a long time, it has been unclear why foehn descends into the valley, and many theories have emerged over the last century (Sprenger et al., 2016). All described different physical processes that eventually happen in the mountain lee. Despite the efforts to find a unified framework to describe the descent, those studies lead to the conclusion that depending on the exact weather situation, different mechanisms must be responsible. Thus, the

[1] The term "meso" usually refers to a scale range from a few kilometers up to 100 kilometers. Everything above is usually referred to as "synoptic".

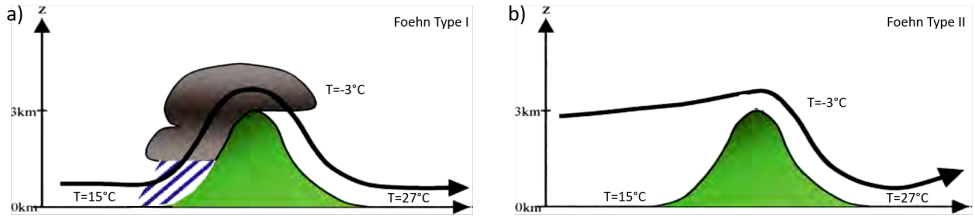


Figure 1.2: a) Shown is a schematic representation of Foehn Type I, which is associated with precipitation and the gain of latent heat on the windward side. b) Shown is a schematic representation of Foehn Type II, which associated with orographic blocking on the windward side. Figures taken from Steinacker (2006).

reductionist view for an all-encompassing foehn descent theory has been largely abandoned (Sprenger et al., 2016). For additional information on the different descent theories, it is referred to Steinacker (2006).

Synoptic causes

Now, we move on to the synoptic situation of foehn. Even though foehn areas are mainly confined to mountain valleys, Richner and Hächler (2013) argued that foehn is a mesoscale phenomenon that derives from a synoptic weather condition (i.e., the weather condition over Europe). Foehn is observed on both sides of the Alps. South foehn refers to foehn on the northern side of the Alps. Similarly, north foehn refers to foehn on the southern side of the Alps (Richner and Hächler, 2013). Since both types are of interest for this work, we briefly explain their respective synoptic situations in the next paragraphs.

The most important condition to produce foehn is a difference in the pressure of air masses on both sides of the Alps (Zweifel et al., 2016). The typical synoptic pressure field present during south foehn can be seen in Figure 1.3. Most of the time, we observe a high-pressure system over Italy and a low-pressure system over the British islands, which cause the necessary (but not sufficient) strong pressure gradient over the Alps. The isobars show their foehn-typical s-curvature (called "foehn-knee"). Usually, a south/south-westerly flow over the Alps is measured. Furthermore, often an additional hydrostatic gradient, caused by a temperature difference between the air masses, is prevalent (Gerstgrasser, 2017). Note that this is only the primary synoptic situation for so-called deep foehn. For example, shallow foehn, where the foehn flow is not necessarily aligned with the flow of the air masses aloft, is driven by a different mechanism. An in-depth summary of different synoptic conditions leading to foehn can be found in Gerstgrasser (2017). For a detailed discussion of a foehn case study and its synoptic development, see Hächler et al. (2011).

Compared to south foehn, north foehn shows precisely the opposite pres-

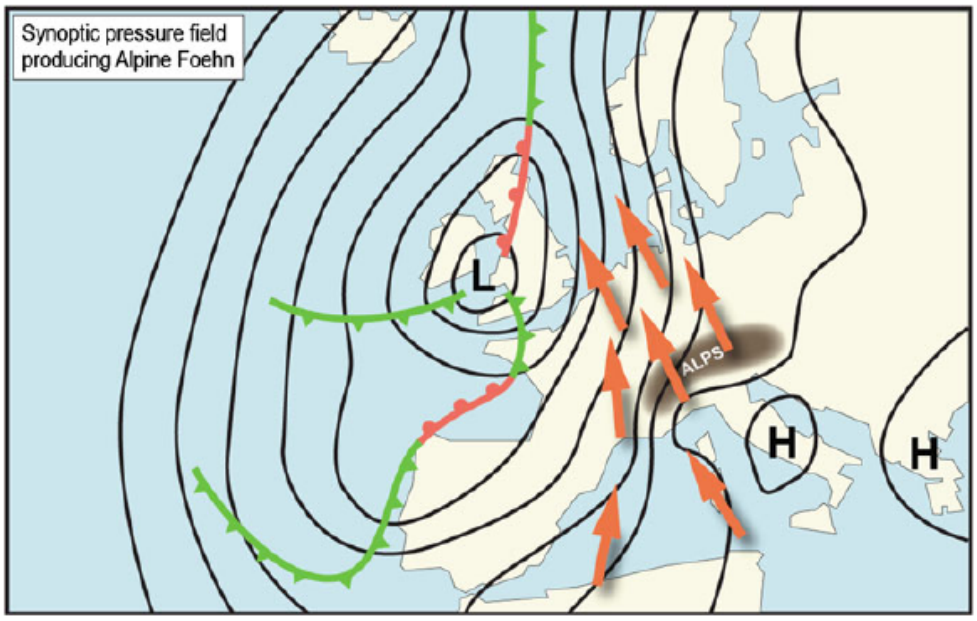


Figure 1.3: Typical synoptic pressure field during south foehn. Figure taken from Richner and Hächler (2013).

sure pattern, i.e., the prevalence of a high-pressure system over the Atlantic/ British Islands and a low-pressure system over south-eastern Europe/Italy (Cetti et al., 2015). Furthermore, due to the break-in of cold air masses to the north of the Alps and the orographic blocking of these by the Alps, a strong hydrostatic gradient develops. For a more detailed north foehn case study, see Kljun et al. (2001).

Due to this synoptic-mesoscale cause-effect relationship, in general, it is possible to deduct foehn from its large-scale fingerprint over Europe (Drechsel and Mayr, 2008). Thus, in the next step, we look at the current state of foehn diagnosis and forecast.

1.1.2 Current state of foehn diagnosis and forecast

In the past, there was a lack of an objective definition of foehn, and often subjective criteria were applied (Zweifel et al., 2016). Thus, in the first section, we introduce different foehn diagnosis techniques, which "nowcast" (compared to forecast) foehn. As a next step, we then outline how foehn is forecasted in the following section. Lastly, based on the former sections, we present what currently is known about the occurrence of foehn from a climatological point of view.

Foehn diagnosis

Measuring foehn has a long tradition in the Alps. In Altdorf, the first measurements date back to 1864, when foehn got diagnosed three times a day by subjectively defined criteria for temperature, humidity, and wind speed (Richner et al., 2014).

Until the mid of the 20th century, a common approach for foehn identification was the (subjective) three-criteria definition introduced by Conrad (1936) and used by Obenland (1956). Here, foehn is defined by abrupt temperature, humidity, and wind speed changes, and the wind direction must adjust along the valley axis (as seen in Figure 1.1). However, as Vergeiner (2004) criticized, this definition is purely phenomenological and does not capture the physical processes in the lee side of the mountain. Furthermore, he argued that this method would cause misjudgments due to the required subjective rating of a human scientist. Mostly, the similar properties of weak foehn events and radiatively driven downslope flows are hard to distinguish even for experts (Plavcan et al., 2014; Vergeiner, 2004).

Gutermann (1970) was the first one to develop an objective and automated classification. In his work, he utilized Fisher's (linear) discriminant analysis to define an objective index based on wind, temperature, and humidity anomalies at the valley station. However, this method still did not rely upon or capture the physical foehn processes which Vergeiner (2004) discussed.

To compensate for the physical shortcomings, Vergeiner (2004) developed an objective foehn classification for the Wipp valley in Austria. In detail, he included criteria for the potential temperature difference $\Delta\theta$ between a crest and the valley station and wind direction dd at both stations. With this, he expanded the work of Schuetz and Steinhauser (1955), which first included the former potential temperature difference. In his summary, Vergeiner (2004) concluded that the results with his technique are comprehensible and reproducible and can potentially be adapted to other foehn locations in the world. However, as Plavcan et al. (2014) annotated, the main draw-back for this method is still having to manually determine a threshold for $\Delta\theta$ for each location individually.

Another approach was developed by Duerr (2008), which allowed to utilize real-time data from the Swiss automatic weather stations for automatic classification of south foehn at various locations all over Switzerland. This method relied on objective thresholds, which were derived from a statistical analysis of ten years of automatic weather station data. These thresholds included the potential temperature difference $\Delta\theta$ between the alpine crest station Gütsch and the valley station, the wind direction DD at Gütsch and the valley station, and relative humidity UU and wind speed FF at the valley station. Since 2008, MeteoSwiss operationalized Duerr's method to auto-

matically classify foehn for every ten-minute interval (Richner and Hächler, 2013). In addition, Cetti et al. (2015) utilized the same method to derive an objective index for north foehn in the Swiss canton Ticino.

Plavcan et al. (2014) aimed to remove the need for manually determining hard thresholds by moving from a deterministic model to a probabilistic method. By utilizing a Gaussian mixture model, they managed to not only derive a foehn index but also assign an uncertainty to each measurement. This method is thought to be location independent, and thus their most simple model is applicable even in the absence of a crest station at the cost of a few wrongly classified observations (Plavcan et al., 2014). They conclude that for the Wipp valley, one year of measurement data sufficed to achieve satisfactory diagnosis results after fitting the model - compared to a decade in Duerr (2008). Later, Plavcan and Mayr (2015) have shown that their method also generalizes well to other locations in the Alpine region.

Lastly, Sprenger et al. (2017) also used a probabilistic approach by leveraging the potential of current state-of-the-art machine learning algorithms. They trained an AdaBoost algorithm on common foehn predictors (for example, pressure differences or wind speeds) to recognize patterns that cause foehn in Altdorf. For this, they used three years of data from the Consortium for Small-Scale Modeling's (COSMO-7) numerical weather prediction (NWP) model. Their target variable, i.e., a binary variable describing whether foehn is prevalent in Altdorf, got derived from Duerr (2008)'s method described above. Finally, with their method, they were able to achieve good scores and showed that it is possible to derive foehn from larger-scale conditions in NWP data. Furthermore, they were especially interested in the most important predictors for the model. Those turned out to be pressure differences over the alps and wind speeds at various measurement stations. Like Plavcan et al. (2014), in principle, their algorithm is also applicable to other locations, given that already a few years of observational foehn data exists.

Reaching beyond the Alpine region, foehn detection is also used in other parts of the world. For example, Laffin et al. (2019) used an XGBoost model to classify different foehn strengths related to the ice melt in Antarctica.

After describing how foehn can be objectively and automatically classified on past and present data, one might now impose the more challenging question of forecasting foehn.

Foehn forecasting

Foehn is traditionally forecasted from the output of NWP models (Zweifel et al., 2016). However, due to the insufficient representation of the topography in NWP models, the direct output of such models has been shown not to accurately represent foehn flows. The horizontal grid resolution of cur-

rent operational NWP models is about one kilometer, and foehn valleys are only coarsely represented. Burri et al. (2007) did show in a case study that the COSMO-2 model overestimates the area affected by foehn and underestimates its wind speeds and temperature. Wilhelm et al. (2012) concluded in his more systematic review of the COSMO model that the foehn frequency is estimated best for Alpine stations but overestimated for fore-alpine and plateau stations. The wind speed of foehn events was found to be underestimated close to the alpine ridge and overestimated further away from it. Lastly, the errors for the wind variables and temperature during foehn were large, which agrees with the findings in Hächler et al. (2011). The relative humidity was represented as the worst of all variables. To sum up, due to the insufficiency of NWP models in accurately capturing foehn, one is still heavily dependent on the skill of experienced meteorologists (Richner and Hächler, 2013).

For this reason, in the past years, model output statistics (MOS) gained more popularity (Glahn and Lowry, 1972). Instead of exclusively looking at the raw output of NWP models, with MOS, one further processes this data by applying statistical techniques to achieve more accurate predictions. The gain in accuracy is achieved by looking at the known synoptic causes for foehn instead of the poorly resembled foehn itself.

Widmer (1966) was the first one to predict foehn in Altdorf. For this purpose, he used Fisher's discriminate analysis to establish an index (nowadays referred to as Widmer index) based upon the forecasted surface level pressure difference and geopotential height difference over the Alps. If this index surpasses a certain threshold, foehn will set in the next 12-36h with a probability of 70%. Later, Courvoisier and Gutermann (1971) were able to further refine and simplify this index. Until today, the Widmer index is used operationally as one of the measures to forecast foehn at MeteoSwiss.

Drechsel and Mayr (2008) analyzed European Centre for Medium-Range Weather Forecasts (ECMWF) model data with MOS. They focused on the cross-alpine pressure difference and potential temperature difference to forecast foehn in the Wipp Valley. They found that, even up to three days, a reliable forecast can be established. Furthermore, they stated that their work is essentially expandable to other locations, as long as the obstacle (i.e., mountain) responsible for foehn is represented at least in a coarse way. The only limitation comes from the existence of observational foehn data. However, with suitably located automated weather stations, this shortcoming can be circumvented.

Lastly, Zweifel et al. (2016) used a standard machine learning technique (namely logistic regression) to forecast foehn in Altdorf 15 and 39 hours in advance upon to data from two NWP models (namely COSMO-2 and ECMWF). Furthermore, they applied the same technique to forecast north

foehn in Piotta. Also, they achieve considerably good results. Again the limitation of this technique comes down to the existence of reliably labeled foehn data.

Note, there have also been efforts in other parts of the world to forecast foehn (although it may not necessarily be called foehn there), which shall briefly be mentioned here. Oard (1993) used a multiple linear regression model to forecast the Chinook in the Rocky Mountains. Mercer et al. (2008) used several machine learning methods (namely linear regression, neural networks, and support vector machines) to forecast downslope windstorms in the Rocky Mountains.

Combining all of the findings above, one can conclude that, in general, it is possible to diagnose and predict foehn from its larger scale fingerprint instead of having to rely on NWP models. In order to verify the results, almost all techniques were tested to reproduce the existing foehn climatology at a given location. Thus, before we move to the final objectives of this part in Subsection 1.1.3, we briefly discuss the climatology of south foehn in Altdorf and north foehn in Lugano.

Climatology

An aggregated foehn time series for Altdorf can be seen in Figure 1.4. Gutermann et al. (2012) showed that there is no long-term trend discernible in the foehn time-series. On average, 483 hours of foehn are observed in Altdorf per year (Gutermann et al., 2012).

Even though the foehn frequency does not show a trend over the years, it yet proves to have an almost constant, strong intra-annual variability. The typical monthly distribution is shown in Figure 1.5. One can observe the annual cycle, with a maximum frequency during spring and a minimum during summer. Gutermann et al. (2012) did show that this pattern is fairly stable across the whole foehn time series, only in the past the frequency of foehn was higher in March than it is today. A mechanism for that could not be identified.

A climatology for several north foehn stations has been created by Cetti et al. (2015). For Lugano, on average, 620 hours of foehn are observed in the analyzed period from 1993 until 2014. Here, no long-term trend could be discovered, too. However, as the south foehn, north foehn shows a pronounced intra-annual cycle shown in Figure 1.6. As we obtain from the figure, the winter/spring months (FMA) show the highest frequency, while the summer/autumn months (ASO) present the lowest frequency.

There have also been efforts in other parts of the world to establish a foehn climatology for the respective region. For example, also within the Alps, Ortner (2010) created a south foehn climatology for Innsbruck. On a more

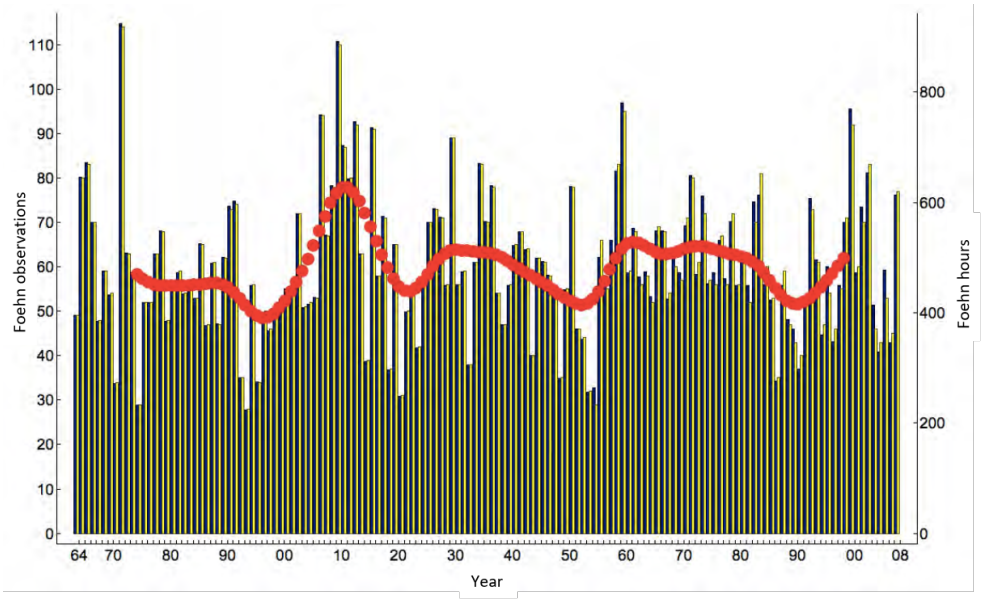


Figure 1.4: Foehn time series from 1864 until 2008 in Altdorf. The left y-axis and the black bars indicate the number of foehn observations each year. Here, one observation is defined as foehn being present at a specific time in the morning, noon, or evening. The right y-axis and the yellow bars display the number of total foehn hours within each year. Note that the foehn hours before 1955 were calculated from the observations as described in Gutermann et al. (2012). The red line marks a Gauss low pass filter with a range of 20 years. Figure taken from Gutermann et al. (2012).

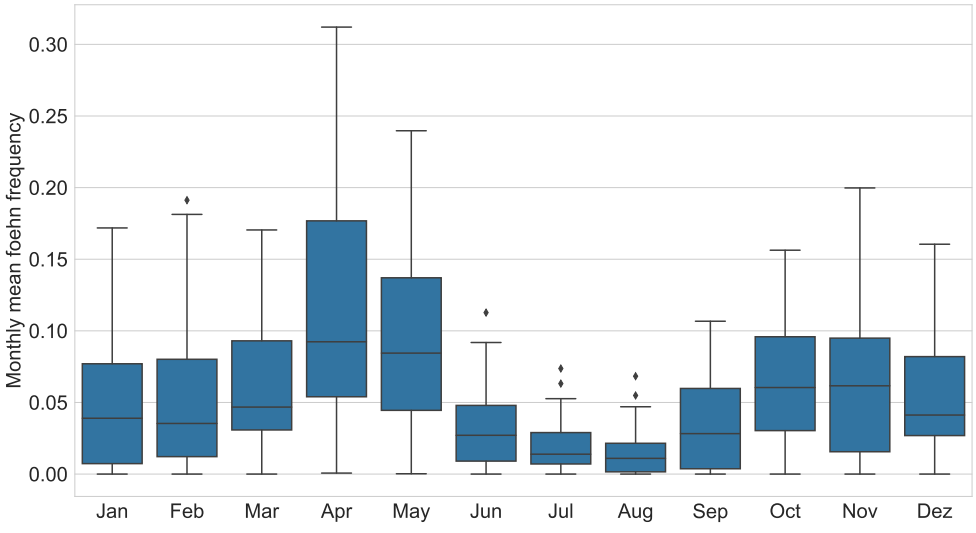


Figure 1.5: Monthly distribution of observed foehn frequency in Altdorf from 1981 until 2019.

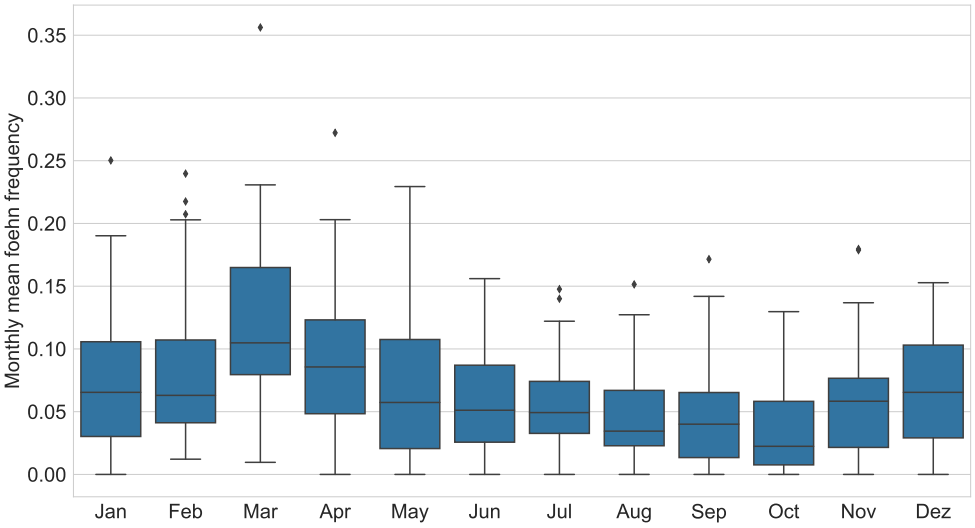


Figure 1.6: Monthly distribution of observed foehn frequency in Lugano from 1983 until 2019.

global scale, Norte (2015) established a climatology for the Zonda winds in the Andes and links it to air quality and wind damage. Cape et al. (2015) investigated a foehn climatology in Antarctica to estimate ice shelf evolution better. Guzman-Morales et al. (2016) calculated the climatology for the foehn-like Santa Ana winds in North America to better understand the influence on the yearly forest fires.

1.1.3 Objectives

Even though no long-term trend seems to be discernible, one might still ask whether the (monthly) foehn frequency would change under a warming climate. We have seen that the synoptic fingerprint of foehn appears to be a good predictor. In the context of this work, we wanted to automatically identify foehn from its large-scale fingerprint from reanalysis data, where observational foehn data exists. Afterward, we wanted to transfer this knowledge to a freely running climate simulation. Subsequently, we compare the occurrence of foehn in simulations of present-day and future climate conditions. Thus, the main questions of this part of the thesis were formulated the following way:

1. Can the synoptic situation which leads to south foehn objectively be identified from coarse NWP reanalysis data? What are the most relevant features on the synoptic-scale?

2. With which level of skill can south foehn be diagnosed from coarse synoptic NWP reanalysis data?

3. Is it possible to generalize the findings from reanalysis data to a freely-running climate simulation?
4. If so, how will the monthly frequency of south foehn differ between present-day climate and a warming future climate?
5. Can we apply the same methodology also for north foehn?

1.2 Impact of foehn on forest fires

Many studies did emphasize the outbreak of fires or severely increased fire damage due to foehn winds. Both, Richner and Hächler (2013) and Sprenger et al. (2016), stated that foehn was responsible for 600 houses burning down in the Swiss canton Glarus in 1861, or, more recently in 2001, for 15 houses in Balzers (Liechtenstein). Carrega (1991) referred back to the two catastrophic fires in France in 1986, where the influence of foehn led to the destruction of several thousands of hectares and several hundred injured people, despite the combined efforts of thousands of firefighters. Zumbrunnen et al. (2009) researched and stressed the importance of foehn winds on forest fires in the Swiss region Valais by comparing foehn and non-foehn areas. For studies relating foehn winds with forest fires also in other parts of the world, see Schroeder et al. (1964), Keeley (2004), Westerling et al. (2004), Sharples et al. (2010), Simpson et al. (2014), and Sharples (2018).

All studies emphasize the importance of foehn for fire outbreak or damage, however, mostly for single events or in a qualitative way. Thus, to our knowledge, no study so far has quantitatively linked foehn winds to forest fires spatially and temporally on the large basis of past fires in the Alps. Actually, Switzerland experiences about 100 forest fires with a total burned area of 300 hectares per year (Pezzatti et al., 2016). Due to this large availability of both foehn wind and forest fire data, Switzerland proves as an ideal study area to determine the quantitative effect of foehn winds on forest fires.

1.2.1 Fire distribution in Switzerland

In Figure 1.7, the spatial distribution of fires in Switzerland is shown. By separating the fires in our fire data set through the main Alpine crest line, approximately 77,4% of all fires are located south of the Alps. Here, the majority is located in the Swiss canton Ticino.

The fires can not only be separated spatially but also temporally into three different fire regimes: winter anthropogenic, summer anthropogenic, and summer natural (De Angelis et al., 2015). Here, the winter anthropogenic fires are confined by the vegetation rest period from December to April and show their maximum during March and April. Note how this coincides with the maximum north foehn occurrence from Figure 1.6. During the winter period, all fire ignitions are due to anthropogenic influence, and the fire events mainly consist of rapidly spreading (surface) fires (De Angelis et al., 2015). In contrast, summer fires instead show a slow-spreading pattern and are of human as well as natural (i.e., lightning) origin. In fact, in July and August, a fire is equally likely to be induced by lightning and humans (Pezzatti et al., 2016). The lighting-induced underground fires occur on a higher elevation and at steeper slopes than anthropogenic fires and thus are more

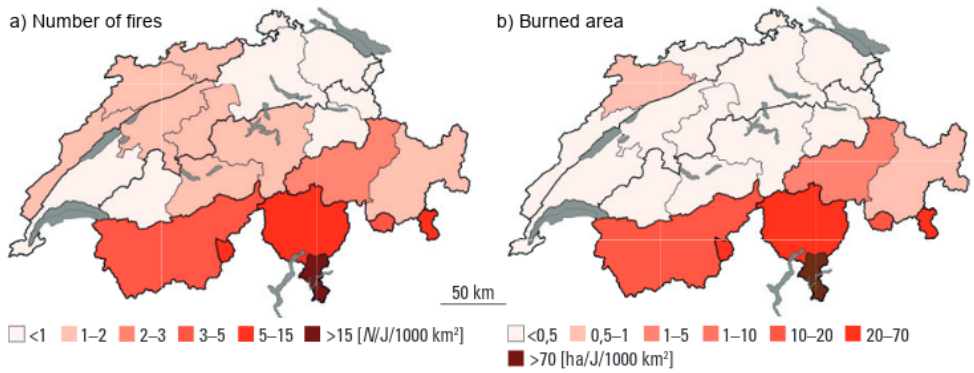


Figure 1.7: Spatial distribution of the number of fires (a) and the burned area (b) of all forest fires within Switzerland from 1990-2014. Figure taken from Pezzatti et al. (2016).

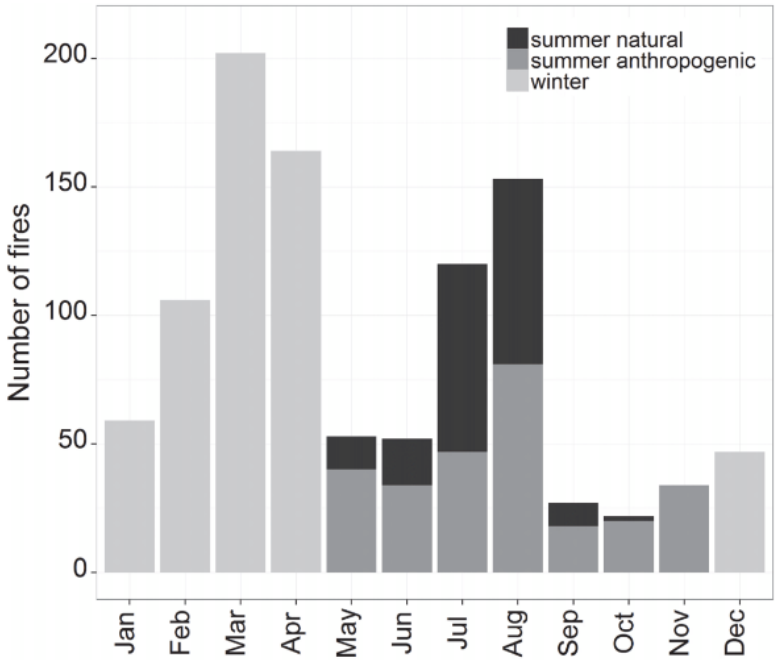


Figure 1.8: Monthly distribution of the number of fires in Ticino, disaggregated by fire regime. The diagram includes fire data from 1991-2012. Figure taken from De Angelis et al. (2015).

challenging to extinguish (Conedera et al., 2006). The monthly distribution of fires, disaggregated by the fire regimes, is shown in Figure 1.8.

1.2.2 Foehn-Fire interaction

As already briefly stated in the introduction, foehn winds can influence forest fires in two different ways. We investigate both mechanisms in this section and will correspondingly formulate the objectives in the next section.

First, we look at the period before a potential fire ignition. Due to its dry and warm nature (i.e., air humidity as low as 20%), foehn can quickly dry out ignitable material (Pezzatti et al., 2016; Sharples et al., 2010; Reinhard et al., 2005). Measures that can be used to assess the potential daily danger of a fire are the so-called Fire Weather Indices (FWIs). Essentially, FWIs are functions that map meteorological variables to the corresponding fire danger and are tuned for a specific type of forest. A common characteristic of all FWIs is that almost all depend on the meteorological variables temperature TT, humidity UU, precipitation P, and wind speed FF as input parameters. The exact composition for all twelve relevant FWIs is nicely summarized in Pezzatti et al. (2016). Just from the definition of a foehn wind, one sees how its characteristics in TT, UU, and FF (see Figure 1.1) are directly linked to the FWIs. Sharples (2018) demonstrates how the onset of a foehn wind in 2008 significantly increased the Forest Fire Danger Index, which is one of the twelve relevant FWIs, by a factor of six. Consequently, we hypothesized that foehn severely improves ignition conditions, thus potentially leading to an increased number of fires.

Second, foehn winds can impact the fire spread when it coincides with the fire ignition or an already burning fire. As Byram (1959) and Carrega (1991) stated, is air movement one of the major fire behavior factors. Wind does affect the burning rate through constant oxygen influx, and fire spread by tilting the flames forward, leading to increased heat advection. Both mechanisms are essential for small fires to build up their intensity (Byram, 1959). By doing so, the onset of foehn can amplify wildfires (Sharples, 2018) and drastically impact the containment of fires and firefighters' safety (Sharples et al., 2010). Moreover, if a fire has not been completely extinguished, according to Sharples et al. (2010), it has significant potential to flare up again. Lastly, the longer it takes to extinguish a fire, the more likely it will be impacted by foehn.

1.2.3 Objectives

In accordance with the last section, we investigated the impact of foehn for two different time frames: before and during a fire. In the former case, we wanted to know how much foehn fosters improved ignition conditions and thus increases the resulting number of fires. In the latter case, we determined how foehn during the ignition period influences the size of a forest fire. The objectives were formulated in the following way:

1. How do foehn winds influence the outbreak of forest fires?
    - Are longer foehn periods associated with an increased number of fires?

- Are warmer foehn winds associated with an increased number of fires?

2. How do foehn winds influence the severity of forest fires?
    - Are longer foehn periods during the ignition phase of a fire associated with a larger burned area?
    - Are higher foehn wind speeds during the ignition period associated with a larger burned area?

Chapter 2

Methodology

In Section 2.1, we first outline our approach to answer the given objectives described in the introduction. In Section 2.2, we explain the different algorithms used during this thesis.

2.1 Conceptual methods

2.1.1 Future development of foehn winds

The basic concept was to train a statistical model (i.e., a machine learning model) on past synoptic weather data to predict (i.e., diagnose) foehn at a given location and make it recognize the distinct synoptic fingerprint for foehn. Here, we interpreted this set-up as a binary classification problem, where we wanted to identify whether foehn is prevalent or not. Then, we employed this model on freely running climate simulations to study the development of foehn. Note that this approach implicitly assumed that the causes for foehn do not vary over the next decades.

For south foehn, we chose Altdorf due to its unique position. The valley of Altdorf is utterly perpendicular to the Alpine crest and a long and homogeneous foehn time series exists (see Subsection 3.1.3). Furthermore, as argued in Gutermann et al. (2012), local effects like cold pools play a smaller role compared to the synoptic situation. Thus, Altdorf was determined as a suitable candidate for our south foehn analysis.

For north foehn, the location of Lugano was chosen. Again, this is due to the perpendicular orientation of the valley. Furthermore, Lugano shows even slightly more foehn events than Altdorf, making it easier to train an algorithm. Besides, as a separate analysis has shown, most forest fires in Switzerland are located in the area of Lugano.

For the purpose of identifying foehn, we trained and tested a model on

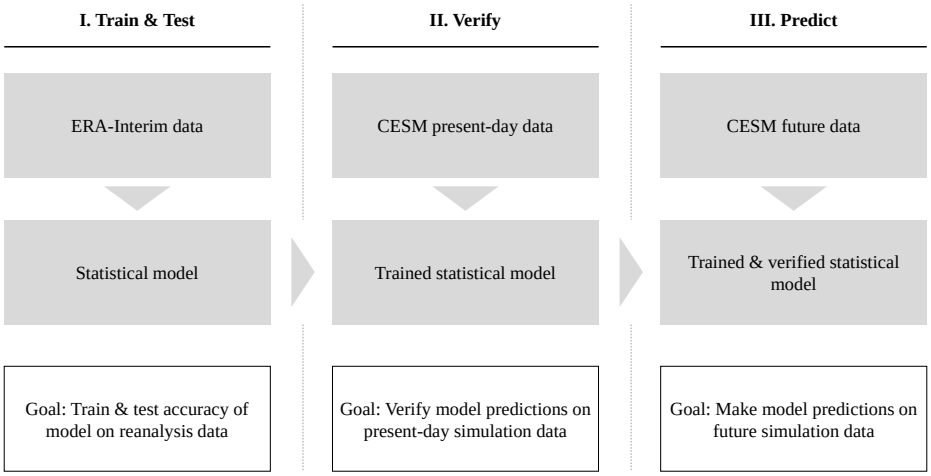


Figure 2.1: Conceptual methodological outline used to evaluate the future development of foehn.

data from the ERA-Interim reanalysis (see Subsection 3.1.1). Afterward, we utilized the model to make predictions on CESM simulation data (see Subsection 3.1.2) of the present-day climate to ensure a viable transition from reanalysis data to a freely running climate simulation. The latter cannot be compared to observational data via coincidence in time; thus, we could not directly train on CESM data. However, to make the model generalize better, in a later approach, we indirectly also utilized the CESM samples in the training procedure. Finally, we used the model to make predictions on CESM simulation data for a warming future climate. Figure 2.1 shows a schematic representation of this procedure. At all steps, we utilized various techniques to learn from the model and tried not to treat it as a black-box. For technical details regarding the implementation, refer to Section 4.1.

2.1.2 Impact of foehn on forest fires

We investigated the objectives for this part from the viewpoint of descriptive statistics. Conceptually, the idea was to spatially map each forest fire to the closest meteorological station and control for foehn presence.

For this aim, we were provided with forest fire data in Switzerland in the past 40 years (for details, see Pezzatti et al. (2010)). Figure 2.2 shows all forest fires for this period and stations where foehn data is available.

Next, we retrieved foehn data from the closest station for the relevant period of each fire. After aggregating, cleansing, and merging this data with the fire data (e.g., the burned area), we were able to make quantitative statements

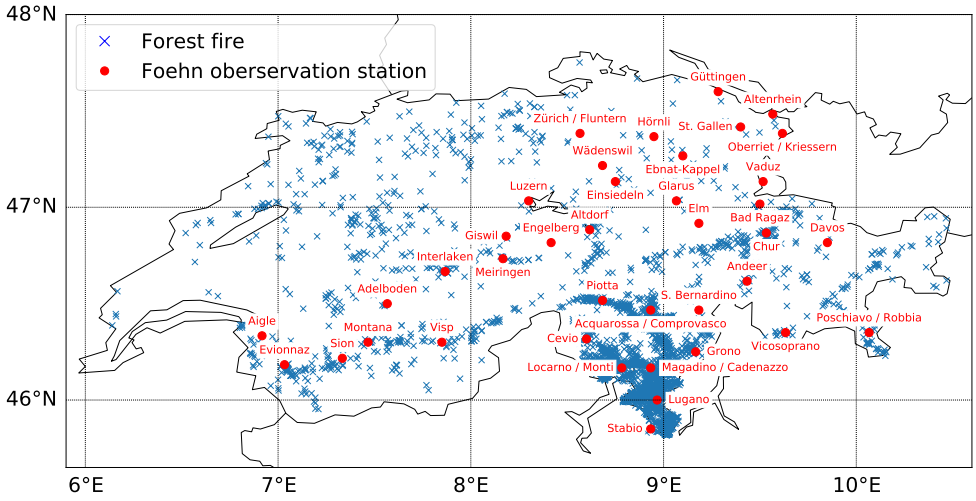


Figure 2.2: Shown are all forest fires from 1980 until 2019 (blue). Meteorological stations where foehn data is available are depicted with their name (red).

about how foehn statistically influenced the number of fires and the burned area. For implementation details, see Section 4.2.

2.2 Machine Learning Algorithms

In this section, we describe the inner workings of the algorithms which we used for our analysis. Since all of them deal with the topic of machine learning, we first introduce some necessary terminology.

Machine learning is often clustered into three different segments: Supervised machine learning, unsupervised machine learning, and reinforcement learning (Mehta et al., 2019). As described above, we approached the problem as a binary classification task and thus dealt with a supervised learning problem. Under such circumstances, especially tree-based models have been shown to yield good performance on structured data sets while requiring little preprocessing and allowing for more interpretability (Mehta et al., 2019; Hastie et al., 2009).

In the next subsections, we describe the range of tree-based algorithms utilized in this thesis. We will slowly increase complexity from simple decision trees in Subsection 2.2.2 over random forests in Subsection 2.2.3 to gradient boosted trees in Subsection 2.2.4. Each time we explain the inner-workings of the algorithm and their advantages and disadvantages. Lastly, we finish by describing several model-agnostic evaluation techniques in Subsection 2.2.5.

2.2.1 Notation

The typical problem set-up for any supervised learning problem presents itself as follows: First, we are given N observations or samples $(\boldsymbol{x}, y)$, which are assembled in a dataset $\boldsymbol{D} = \{(\boldsymbol{x}^{(i)}, y^{(i)})\}_{i=1}^{N}$. Hereby, each sample vector $\boldsymbol{x}^{(i)} = (x_1, ..., x_M) \in \mathbb{R}^M$ combines M different features x_m. The sample vector represents one point of the feature space $\boldsymbol{X} = (X_1, ..., X_M)$, where the X_m denote random variables which are distributed with a generally unknown probability. Hence, $\boldsymbol{x}$ is one possible realization within $\boldsymbol{X}$. Next, each observation $(\boldsymbol{x}, y)$ usually also includes an (observed) target label $y \in \mathbb{R}$. From now on we will use the indexed version $(\boldsymbol{x^{(i)}}, y^{(i)})$ whenever we talk about an observation in $\boldsymbol{D}$ and $(\boldsymbol{x}, y)$ when we talk about a general observation (which could be in $\boldsymbol{D}$ but also new data).

The goal of the model is to find (i.e., fit) a function $f(\boldsymbol{x}; \boldsymbol{\theta}) : \mathbb{R}^M \rightarrow \mathbb{R}$ on $\boldsymbol{D}$ in such a way, that

$$y = \hat{y} = f(\boldsymbol{x}; \boldsymbol{\theta}) \tag{2.1}$$

Hereby, $\boldsymbol{\theta}$ represents the weights or parameters of f, which are determined by fitting the model on $\boldsymbol{D}$. When we write $\hat{y}$, we do mean the predicted value f yields for $\boldsymbol{x}$.

For the fitting process, we need a scalar metric which tells us how well our model is already approximating the observed label. Usually, this metric is called the loss function $l(y, \hat{y}) = l(y, f(\boldsymbol{x}; \boldsymbol{\theta}))$ and measures the deviation of the prediction $\hat{y} = f(\boldsymbol{x}; \boldsymbol{\theta})$ from the observed label y. We can now find the optimal weights $\boldsymbol{\theta}^*$ by minimizing the loss function for all observations in $\boldsymbol{D}$

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} l(y^{(i)}, f(\boldsymbol{x}^{(i)}; \boldsymbol{\theta})) \tag{2.2}$$

This problem set-up is very general and can be applied to a wide range of problems. For example, $\boldsymbol{X}$ could represent various variables describing the state of the atmosphere (e.g., pressure or wind flow). Consequently, $\boldsymbol{x}$ would describe one observation of those variables at a given point in time. In accordance, the target label y could determine whether foehn is prevalent in Altdorf. For binary classification, the only restriction on this general setup is that $y \in \{0, 1\}$. This means, that y tells us whether $\boldsymbol{x}$ belongs to class 0 ($y = 0$; no foehn) or class 1 ($y = 1$; foehn).

Before training the model, $\boldsymbol{D}$ is split up into a training dataset $\boldsymbol{D}_{train}$ and a test dataset $\boldsymbol{D}_{test}$. The idea is to optimize the parameters $\boldsymbol{\theta}$ on $\boldsymbol{D}_{train}$ (often referred to as "training"). Afterward, the model's performance is evaluated on the before unseen $\boldsymbol{D}_{test}$ (referred to as "testing"). This is done to avoid overfitting and to ensure transportability of results. Several techniques for

the model evaluation will be explained in Subsection 2.2.5. One often trains the model on randomly selected 80% of the data and then evaluates it on the remaining 20% (Mehta et al., 2019). This usually serves as a good first indication and can be adjusted to the problem at hand.

In the next steps, we will put the above theory into a more detailed context by explaining various models used for the analysis.

2.2.2 Decision Trees

A decision tree is a very simple machine learning model often used for regression or classification (Louppe, 2014; Hastie et al., 2009). Note that from here on, we mainly deal with the concept of two-class classification trees since that is what we use in this thesis. However, in the end, we also describe the necessary modifications to turn a classification tree into a regression tree since we indirectly utilize them in gradient boosted trees later. Multi-class classification with classification trees is a straightforward generalization, and an excellent description can be found in Hastie et al. (2009).

Classification trees

A classification tree works by partitioning the observations $(\boldsymbol{x}^{(i)}, y^{(i)}) \in \boldsymbol{D}_{train}$ into J mutually disjoint regions $R_1, ..., R_J \subset \boldsymbol{X}$ and assigning a value p_j to each region R_j. In case of binary classification, p_j denotes the probability of a $\boldsymbol{x} \in R_j$ belonging to class 1. Likewise, with probability $1 - p_j$ it belongs to class 0. Mathematically, we can write a classification tree as

$$f^{CT}(\boldsymbol{x}; \boldsymbol{\theta}) = \sum_{j=1}^{J} p_j I(\boldsymbol{x} \in R_j) \tag{2.3}$$

where $\boldsymbol{\theta} = \{R_j, p_j\}_{j=1}^{J}$ resembles the weights and

$$I = \begin{cases} 1 & \text{if } \boldsymbol{x} \in R_j \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

is the indicator function. After one has decided upon the partition regions R_j, the calculation of p_j is naturally determined by

$$p_j = \frac{1}{N_j} \sum_{\boldsymbol{x}^{(i)} \in R_j} y^{(i)} \tag{2.5}$$

where N_j is the sum of observations in R_j. We calculate the final prediction $\hat{y}$ by controlling if the result of Equation 2.3 is larger than a certain threshold

$c \in [0,1]$. We define

$$\hat{y} = \Theta(f^{CT}(\boldsymbol{x};\boldsymbol{\theta}) - c) = \begin{cases} 1 & \text{if } f^{CT}(\boldsymbol{x};\boldsymbol{\theta}) \geq c \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

with the Heaviside function Θ. The threshold c can be set manually and differs between applications. Instead of manually choosing a threshold, we could have also predicted the more likely class, which corresponds to a fixed threshold of $c = 0.5$. However, Equation 2.6 gives us more flexibility in tuning an already trained model in hindsight, depending on the metric of interest.

Now that we have explained how a classification tree makes a prediction for already given R_j's, we are left with the more difficult part, which is how we can calculate the R_j's from $\boldsymbol{D}_{train}$.

Classification trees partition $\boldsymbol{D}_{train}$ by recursively splitting $\boldsymbol{X}$ into a set of several rectangular subregions R_j in a binary way (Hastie et al., 2009). Let us consider a tree which only performs one split. It divides $\boldsymbol{X}$ into two subregions R_1 and R_2, by splitting feature m at a split point s

$$R_1(m,s) = \{X|X_m \leq s\} \quad \text{and} \quad R_2(m,s) = \{X|X_m > s\}. \tag{2.7}$$

where we have made explicit the dependence of the R_j on the corresponding feature m and split-point s. It is obvious, that we want to find the optimal values for m^* and s^* in order to achieve better discrimination between classes in R_1 and R_2 than in X. We can do so by calculating

$$m^*, s^* = \arg\min_{m,s} \left[N_1 \underbrace{2p_1(1-p_1)}_{Q(p_1)} + N_2 \underbrace{2p_2(1-p_2)}_{Q(p_2)} \right], \tag{2.8}$$

where

$$Q_j = Q(p_j) = 2p_j(1-p_j) \tag{2.9}$$

denotes the so-called "Gini impurity" for two classes. N_j weights the impurity according to the amount of observations in R_j. In general, the Gini impurity Q measures the probability of an observation being miss-classified within R_j if it would be labeled randomly with probability p_j. Q is a concave function and as one can easily obtain from Equation 2.5, it becomes zero, if and only if there is only one class present in R_j (i.e., if R_j is "pure"). Consequently, for a good separation between classes, we want to minimize Q for both subregions R_1 and R_2 weighted by their corresponding N_j. Note how Q corresponds closely to the loss function mentioned above, and $\sum_j N_j Q_j$ is the objective we want to minimize to obtain $\boldsymbol{\theta}^*$ (compare Equation 2.2).

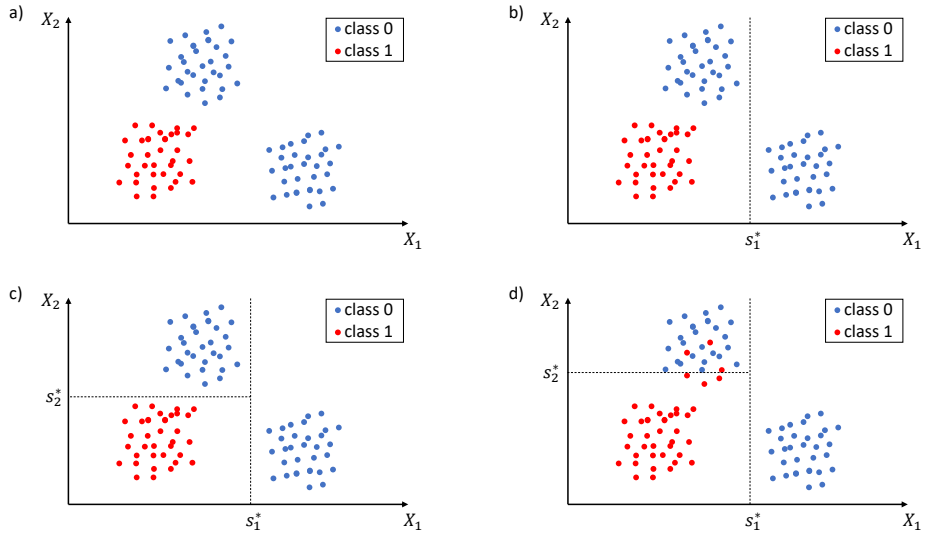


Figure 2.3: Shown is a graphical interpretation of fitting a classification tree to data. a) Plotted is an exemplary dataset $\boldsymbol{D}$ with two features X_1 and X_2. Each dot represents an observation $\boldsymbol{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}) \in \boldsymbol{D}$. The blue color indicates that the sample belongs to class 0 (e.g., no foehn), the red color indicates class 1 (e.g., foehn). b) First split of the data, where s_1^* and $m_1^* = 1$ are determined by minimizing Equation 2.8. c) Second split on X_2 in the subspace obtained from the first split. The values of s_2^* and $m_2^* = 2$ are again determined by Equation 2.8. d) If the data is not separable under the given features, the model will still be able to determine optimal values for s_j^* and m_j^*, however, for the price of miss-classifications.

Computationally, one can determine s^* easily for each feature m. Thus, by scanning through all inputs, the best values for m and s can quickly be found (Hastie et al., 2009). This is also called the "greedy" algorithm since it considers all possible choices.

Usually, one does not perform a single split (which results in a so-called "decision-stump"), but recursively repeats the procedure described in Equation 2.7 and Equation 2.8 with the new subsets which are generated in each split. This training procedure can also be illustrated graphically (see Figure 2.3).

However, repeating the training procedure infinitely would result in a heavily overfitted algorithm[1]. Hence, one would stop the recursion on a region R_j as soon as

- $Q(p_j) = 0$ or
- $d = d_{max}$ or
- $N_j \leq N_{min}$.

[1]Consider the case where J equals $N - 1$. Here, each $x^{(i)}$ falls into a different R_j.

Hereby, d describes the depth (i.e., amount of consecutive splits on $\boldsymbol{D}_{train}$ to reach R_j) of the tree and d_{max} and N_{min} mark certain fixed thresholds (often called hyperparameters). An pseudo-code implementation of the training procedure can be seen in Algorithm 1.

Lastly, not only the fitting procedure but also the whole tree can be visualized for easier interpretation (see Figure 2.4). Often, the splits are referred to as "Nodes". The ends of a tree which are associated with a prediction via Equation 2.3 and Equation 2.6 are called "Leaves".

Algorithm 1: Classification tree algorithm for two classes

Training

1. Set $R_1 = \boldsymbol{X}$
2. While $(Q(p_j) \neq 0) \wedge (N_j > N_{min}) \wedge (d < d_{max})$ for a R_j:
   a) Calculate optimal split parameters m^* and s^* for this R_j according to Equation 2.8 and split the node
   b) Update R_j's and p_j's
3. Output $\boldsymbol{\theta}^* = \{R_j, p_j\}_{j=1}^J$

Prediction

1. Calculate $f^{CT}(\boldsymbol{x}; \boldsymbol{\theta}^*) = \sum_{j=1}^J p_j I(\boldsymbol{x} \in R_j)$
2. Calculate final prediction via $\hat{y} = \Theta(f^{CT}(\boldsymbol{x}; \boldsymbol{\theta}^*) - c)$

Regression trees

Regression trees work almost in the same manner as classification trees. However, y is now no longer a binary variable, but a continuous one. We can write a regression tree by

$$f^{RT}(\boldsymbol{x}; \boldsymbol{\theta}) = \sum_{j=1}^{J} c_j I(\boldsymbol{x} \in R_j) \tag{2.10}$$

where $\boldsymbol{\theta} = \{R_j, c_j\}_{j=1}^J$ again resembles the weights. In a similar fashion to classification trees, the calculation of c_j given R_j is determined by

$$c_j = \frac{1}{N_j} \sum_{\boldsymbol{x}^{(i)} \in R_j} y^{(i)}, \tag{2.11}$$

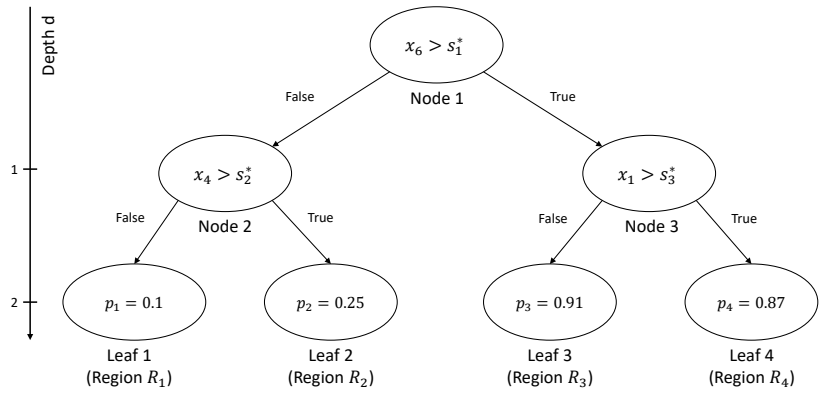


Figure 2.4: Shown is an exemplary structure of an already grown binary two-class classification tree of depth two. At every "Node" it is being decided whether feature x_m of an observation is greater than the learned value s_n^* (n numbering the nodes). As soon as the algorithm reaches a "Leaf", the predicted class $\hat{y}$ for this observation $\boldsymbol{x}$ will be determined according to Equation 2.6 from the previously calculated p_j.

which is simply the mean over all target labels within a region R_j. Since y is a continuous variable, a threshold becomes redundant and the final prediction of the regression tree as given by

$$\hat{y} = f^{RT}(\boldsymbol{x}; \boldsymbol{\theta}^*) \tag{2.12}$$

Thus, for every $\boldsymbol{x} \in R_j$ the regression tree does nothing but predicting the mean c_j, which got learned during training. The training follows the same procedure described above, we only need to replace Equation 2.8 with

$$m^*, s^* = \arg\min_{m,s} \left[c_1(m,s) + c_2(m,s) \right], \tag{2.13}$$

as it is derived in Hastie et al. (2009) assuming a mean squared loss.

Feature importance in a decision tree

Instead of manually inspecting a decision tree, one can utilize the information available about the splits and their contribution to the final prediction in order to construct a single-value metric called feature importance. This allows to compare single features against each other. Breiman et al. (1984) originally proposed to use

$$\mathcal{T}_m = \sum_{t=1}^{J-1} \tau_t^2 I(m(t) = m), \tag{2.14}$$

where the sum goes over all $J-1$ internal nodes. $m(t)$ returns the optimal feature m^* which was identified for the split in node t and I is again an indicator function. τ_t measures the squared weighted reduction in error in node t which was achieved with the identified optimal split (see Equation 2.8). For two-class classification, τ_t can be written as

$$\tau_t = N_t Q(p_t) - (N_{t+1} Q(p_{t+1}) + N_{t+2} Q(p_{t+2})) \tag{2.15}$$

In words, τ_t tells us by how much we can reduce the impurity $Q(p_t)$ in node t by splitting it into the two child nodes which correspond to the impurities $Q(p_{t+1})$ and $Q(p_{t+2})$. Afterward, the formula weights the impurities by their corresponding amount of samples N_{t+1} and N_{t+2}. Equation 2.14 then simply attributes this impurity reduction to each feature m and aggregates it over the whole tree.

For a regression tree Equation 2.15 becomes

$$\tau_t = N_{t+1}(c_t - c_{t+1})^2 + N_{t+2}(c_t - c_{t+2})^2 \tag{2.16}$$

as one can show in a rather lengthy derivation assuming again mean squared loss.

Advantages and disadvantages of decision trees

Advantages In the following, we want to outline the benefits of using decision trees.

- **Interpretability.** A decision tree is straightforward to read, and predictions can be made transparent by simply tracing the decision path (compare Figure 2.4). Decision trees are often referred to as white-box models and have a logic easily understandable even by non-statistically advanced users (Louppe, 2014).
- **Fast inference.** For the classification of a new $\boldsymbol{x}$, a maximum time of $o(d)$ is required. That is especially helpful when there are many observations to classify. Furthermore, they are also relatively fast to construct, with a training time of order $o(MN\log(N))$ for a fully-grown tree (Hastie et al., 2009).
- **Immune to correlated/irrelevant features.** Due to their design, decision trees automatically select the features with the most predictive power (Louppe, 2014; Hastie et al., 2009) and will ignore highly-correlated or irrelevant features after that (since those do not contain additional information).
- **Very little data-preprocessing**. They do not require feature normalization/scaling (in fact, they are invariant under strict monotone transformations) or treatment of outliers. Furthermore, missing training data

does not strictly hinder the training. Thus, according to Hastie et al. (2009) decision trees come closest to an "off-the-shelf" procedure for data-mining.

Disadvantages Before leaving this section, having presented the various advantages of decision trees, we will now discuss their limitations. By doing so, we motivate the next section, which deals with a possible extension of decision trees: Random Forests.

- **Overfitting.** Decision trees easily overfit and generalize bad to new unseen data. Recall that for this reason, decision trees are usually regularized by setting d_{max} or N_{min}.
- **Only single feature splits.** If the data is not oriented in the right way, a decision tree will need many splits to capture even simple relationships (e.g., linear) in the data. Removing this shortcoming and allowing linear combinations of splits can improve performance, however, it hurts interpretability (Hastie et al., 2009).
- **Bias towards high cardinality categorical features**. It is argued in Hastie et al. (2009) that the partitioning algorithm tends to favor categorical features with a high cardinality (i.e., a high number of possible categories). Thus, it is recommended to discard such features.
- **Robustness.** It has been shown that decision trees are susceptible to details in the dataset, and little changes in $\boldsymbol{D}_{train}$ will lead to a highly different structure of the tree (Mehta et al., 2019; Hastie et al., 2009).
- **Accuracy.** Lastly, perhaps the largest disadvantage is, as Hastie et al. (2009) states, that unfortunately, a decision tree seldom achieves an accuracy that is comparable with other machine learning techniques. However, by combining trees in an ensemble (through bagging or boosting), this accuracy can often dramatically be improved (at the cost of interpretability).

2.2.3 Random Forests

Random forests are a more sophisticated algorithm than normal decision trees and were first proposed in Breiman (2001). They work by combining several trees together through a procedure called bagging (which we will discuss below), and further improving accuracy through randomization of the trees. We can write a binary classification random forest with B trees as

$$f_B^{RF}(\boldsymbol{x}) = \frac{1}{B}\sum_{b=1}^{B} f_b^{CT}(\boldsymbol{x};\boldsymbol{\theta}_b) \tag{2.17}$$

where $f_b^{CT}(\boldsymbol{x}; \boldsymbol{\theta}_b)$ is the prediction of the b-th classification tree as in Equation 2.3. Again, in order to keep flexibility, the final prediction is determined by passing Equation 2.17 through the shifted Heaviside function

$$\hat{y} = \Theta(f_B^{RF}(\boldsymbol{x}) - c) = \begin{cases} 1 & \text{if } f_B^{RF}(\boldsymbol{x}) \geq c \\ 0 & \text{otherwise.} \end{cases} \tag{2.18}$$

As one can imagine, growing several trees on the same data set yields the same tree multiple times and does not improve prediction. Thus, a procedure called "bagging" (which is short for "bootstrap aggregation") is used for growing distinct trees. In bagging, we fit each tree not with $\boldsymbol{D}_{train}$, but a bootstrapped version $\boldsymbol{D}'_{train} \subseteq \boldsymbol{D}_{train}$ (i.e., a randomly, with-replacement sampled subset of same size) and then combine the trees in an ensemble. In general, bagging can be used for any low-bias, high-variance (see below) estimator and not only for decision trees.

To further improve the predictive power of a random forest, the resulting trees should be decorrelated as much as possible (Hastie et al., 2009). As Breiman (2001) has shown, an effective way of further improving decorrelation is by considering a random subset of M' features at each split. For classification, usually one would choose $M' = \sqrt{M}$ (Hastie et al., 2009). A pseudo-code implementation can be found in Algorithm 2. In the next step, we examine an explanation for the improvement the described procedure yields. For that, we introduce the bias-variance-decomposition of random forests.

In machine learning, the generalization error E_{out} (i.e., the error on the test set) of the model can be written in the form of a bias-variance-decomposition (Mehta et al., 2019)

$$E_{out} = Bias^2 + Variance + Noise$$

The *Bias* measures the deviation of the expectation value of the model (in the infinite data limit) from the true value. *Variance* measures how much the model fluctuates due to finite sample effects. The *Noise* is due to intrinsic noise in the data and can not be overcome by any statistical model. Since we want to minimize E_{out}, we can reduce either the *Bias* or the *Variance*. For bagging, it can be shown that the *Bias*-Term in the ensemble of classifiers is the same as the bias of any individual classifier (Hastie et al., 2009; Mehta et al., 2019). Thus, the benefit a random forest brings over a decision tree stems solely from reducing *Variance*. Mehta et al. (2019) shows that for bagging the *Variance* can be written as

$$Variance = \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2, \tag{2.19}$$

Algorithm 2: Random forest algorithm for two-class classification (adopted from Hastie et al. (2009))

Training

1. For b=1 to B:
   a) Draw a bootstrapped sample $\boldsymbol{D}'_{train}$ of size N from $\boldsymbol{D}_{train}$.
   b) Grow a classification tree f_b^{CT} on $\boldsymbol{D}'_{train}$, by recursively repeating the following steps for each node of the tree, until stopping criteria is reached
      i. Select random $M' < M$ features.
      ii. Calculate optimal m^* and s^* for the split.
      iii. Split the node into two child nodes.
2. Output $\{f_b^{CT}\}_{b=1}^{B}$

Prediction

1. Calculate $f_B^{RF}(\boldsymbol{x}) = \frac{1}{B}\sum_{b=1}^{B} f_b^{CT}(\boldsymbol{x}; \boldsymbol{\theta}_b^*)$
2. Calculate final prediction via $\hat{y} = \Theta(f_B^{RF}(\boldsymbol{x}) - c)$

where σ^2 denotes the variance of a single classifier and ρ the positive pairwise correlation between classifiers in the ensemble. By increasing the number of classifiers B in the ensemble (i.e., the amount of trees in the forest), we can bring down the second term. The benefit a random forest has over an ensemble of bagged decision trees is, as already described above, through reducing ρ by only splitting the data on a random subset of features.

As one can imagine, even though we are able to reduce E_{out}, it comes at the cost of interpretability. Especially for large B, it may not be clear how the random forest arrives at its final prediction. Thus, in the next section, we investigate how we can aggregate the feature importance of single decision trees for random forests.

Feature importance of random forests

As we have seen in Equation 2.14, in a decision tree we can attribute the improvement in the split-criterion to a corresponding feature. We can now generalise this equation to random forests by calculating the feature importance $\mathcal{T}_m(f_b^{CT})$ for each of the B trees in the ensemble and then averaging

over the results by

$$\mathcal{T}_m^{RF} = \frac{1}{B} \sum_{b=1}^{B} \mathcal{T}_m(f_b^{CT}) \tag{2.20}$$

Naturally, $\mathcal{T}_m^{RF}$ now gives us the feature importance for the m-th feature determined by the random forest.

Advantages and disadvantages of random forests

Advantages Random forests inherit many advantages which stem from the design of decision trees discussed above. Furthermore, they build on some disadvantages and substantially improve them.

- **Accuracy & Robustness.** The predictions of random forests are more accurate and robust than for single decision trees and are comparable to that of boosted trees (Hastie et al., 2009).
- **Low complexity.** The hyperparameters d_{max} and N_{min} are the same as for all trees in the forest; only B and M' are additional parameters that can be tuned.
- **Overfitting is seldom.** Breiman (2001) argues, that a random forest is immune to overfitting, when increasing B. Hastie et al. (2009) counters that, in general, overfitting is possible with fully grown trees. However, this occurs very seldom and at minimal cost of accuracy.
- **Parallelizability.** Since the trees in a random forest are independent from another, a random forest algorithm is highly parallelizable.
- **Usage of the OOB score.** One can calculate the so-called out-of-bag (OOB) score, which uses the observations not included while growing a tree due to the bagging procedure. That eliminates the need for setting aside a test set. The OOB score is almost identical to that obtained by cross-validation (Hastie et al., 2009).

Disadvantages In contrast, some advantages of using a decision tree are lost for random forests.

- **Resource consumption.** Although random forests are parallelizable, they still require larger computational resources for the training if $BM' > M$. Since all trees need to be evaluated for a prediction, inference time is with $o(Bd)$ also longer for non-parallelized code.
- **Lower interpretability.** Perhaps the greatest drawback compared to decision trees is the loss of interpretability. As Hastie et al. (2009) states, a bagged decision tree is no longer a tree.

2.2.4 Gradient Boosted Trees & XGBoost

Another method to combine trees into an ensemble is called boosting. In contrast to bagging, where the ensemble members are build independently from each other, in boosting, we build them sequentially. Here, each classifier is trained to reduce the errors of the previous ensemble members. This method was first implemented by Friedman (2001) in gradient boosted trees. The computational performance is pushed to its limits by Chen and Guestrin (2016) in their more sophisticated implementation named Extreme Gradient Boosting (XGBoost). First, we will explain the mechanism behind gradient boosted trees and then move on to the benefits of XGBoost.

In a very similar manner to Equation 2.17, we construct the boosted tree ensemble by

$$f_B^{BT}(\boldsymbol{x}) = \sum_{b=1}^{B} f_b^{RT}(\boldsymbol{x}; \boldsymbol{\theta}_b). \tag{2.21}$$

Here, the $f_b^{RT}(\boldsymbol{x}; \boldsymbol{\theta}_b)$ are regression trees due to the sequential architecture of the algorithm, even for classification tasks. To convert this into a probability, the logistic function

$$p(z) = \frac{1}{1 + \exp(-z)} \tag{2.22}$$

is used. Then, the final prediction is again determined via the threshold c as in Equation 2.6 by

$$\hat{y} = \Theta(p(f_B^{BT}(\boldsymbol{x})) - c) = \begin{cases} 1 & \text{if } p(f_B^{BT}(\boldsymbol{x})) \geq c \\ 0 & \text{otherwise.} \end{cases} \tag{2.23}$$

Thus, the goal is to maximize $f_B^{BT}(\boldsymbol{x})$ for $y^{(i)} = 1$ and minimize it for $y^{(i)} = 0$ and map it to its corresponding probability afterwards. We can do so by minimizing the loss function l. One possible choice for binary classification is the (negative) logistic loss or binary cross-entropy

$$l_{LL}(y^{(i)}, p^{(i)}) = -(y^{(i)} \log(p^{(i)}) + (1 - y^{(i)}) \log(1 - p^{(i)})). \tag{2.24}$$

We calculate the sum over l on $\boldsymbol{D}_{train}$ for all $f_{b-1}^{BT}(\boldsymbol{x}^{(i)})$ and then add the b-th tree to the ensemble in such a way that it further minimizes the loss. Formally, we can write the determination of $\boldsymbol{\theta}_b^*$ by

$$\boldsymbol{\theta}_b^* = \arg\min_{\boldsymbol{\theta}_b} \sum_{i=1}^{N} l\left(y^{(i)}, f_{b-1}^{BT}(\boldsymbol{x}^{(i)}) + f_b^{RT}(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}_b)\right) \tag{2.25}$$

Usually, l is minimized by gradient descent, hence the name gradient boosting. For the logistic loss, the negative gradient for each observation in $\boldsymbol{D}_{train}$ can be written as residual

$$-\left[\frac{\partial l_{LL}(y^{(i)}, f_{b-1}^{BT}(\boldsymbol{x}^{(i)}))}{\partial f_{b-1}^{BT}(\boldsymbol{x}^{(i)})}\right] = y^{(i)} - p(f_{b-1}^{BT}(\boldsymbol{x}^{(i)})) = r_b^{(i)}. \tag{2.26}$$

The residuals measure the errors of $f_{b-1}^{BT}(\boldsymbol{x}^{(i)})$. Now, we fit the b-th decision tree $f_b^{RT}(\boldsymbol{x}^{(i)}; \boldsymbol{\theta})$ with $r_b^{(i)}$ as the target label and hereby determine the regions $\{R_{jb}\}_{j=1}^{J_b}$. Afterward, the optimal weights for each region are calculated through

$$c_{jb} = \frac{\sum_{x^{(i)} \in R_{jb}} r_b^{(i)}}{\sum_{x^{(i)} \in R_{jb}} |r_b^{(i)}|(1 - |r_b^{(i)}|)} \tag{2.27}$$

as it is shown in Friedman (2001). With those parameters we can construct the b-th tree $f_b^{RT}(\boldsymbol{x}; \boldsymbol{\theta}_b)$ with $\boldsymbol{\theta}_b = \{c_{jb}, R_{jb}\}_{j=1}^{J_b}$. Finally, we add this tree to the ensemble and repeat the procedure until we reach B trees. A pseudo-code implementation of this can be found in Algorithm 3.

Next, in the same manner as we generalized the concept of bagged trees to random forests, we will now build upon gradient boosted trees and (briefly) introduce XGBoost.

XGBoost (Chen and Guestrin, 2016) is an open-source library that implements gradient boosted trees in a computationally highly-efficient way. The XGBoost model has been used with high success in many past applications achieving state-of-the-art performance (Chen and Guestrin, 2016). In physics, the XGBoost model was used to detect signals from Higgs boson related events from vasts amount of data (Chen and He, 2015). Also, XGBoost has already been applied to classify foehn strength in the Antarctic (Laffin et al., 2019).

XGBoost performs better due to algorithmic and system optimizations of gradient boosted trees. We will briefly summarize some of the main improvements XGBoost offers over other gradient boosted trees libraries. For more details, the reader is referred to Chen and Guestrin (2016).

- **Regularization.** The loss function is expanded to include regularization terms, forcing the model to find an accuracy vs. simplicity tradeoff and hence allowing to minimize overfitting. Here, XGBoost builds on the work of Johnson and Zhang (2013).
- **Partially parallelized.** Although the building of sequential trees cannot be parallelized, XGBoost parallelizes the splitting of nodes within a tree.

Algorithm 3: Gradient boosted algorithm for two-class classification (adopted with variations from Friedman (2001) and Hastie et al. (2009))

Training

1. Initialize $f_0^{BT}(\boldsymbol{x}^{(i)}) = 0$
2. For b=1 to B:
   a) Set $p_b^{(i)} = p(f_{b-1}^{BT}(\boldsymbol{x}^{(i)})) = \frac{1}{1+\exp\left(-f_{b-1}^{BT}(\boldsymbol{x}^{(i)})\right)}$.
   b) Compute residuals $r_b^{(i)} = y^{(i)} - p_b^{(i)}$.
   c) Fit a regression tree to $r_b^{(i)}$ and obtain the regions $\{R_{jb}\}_{j=1}^{J_b}$
   d) Compute $c_{jb} = \frac{\sum_{x^{(i)} \in R_{jb}} r_b^{(i)}}{\sum_{x^{(i)} \in R_{jb}} |r_b^{(i)}|(1-|r_b^{(i)}|)}$.
   e) Construct $f_b^{RT}(\boldsymbol{x}; \boldsymbol{\theta}_b^*)$ with $\boldsymbol{\theta}_b^* = \{c_{jb}, R_{jb}\}_{j=1}^{J_b}$
   f) Update $f_b^{BT}(\boldsymbol{x}) = f_{b-1}^{BT}(\boldsymbol{x}) + f_b^{RT}(\boldsymbol{x}; \boldsymbol{\theta}_b^*)$
3. Output $\{f_b^{RT}(\boldsymbol{x}; \boldsymbol{\theta}_b^*)\}_{b=1}^{B}$

Prediction

1. Calculate $f_B^{BT}(\boldsymbol{x}) = \sum_{b=1}^{B} f_b^{RT}(\boldsymbol{x}; \boldsymbol{\theta}_b^*)$
2. Calculate final prediction via $\hat{y} = \Theta(p(f_B^{BT}(\boldsymbol{x})) - c)$

- **Early stopping.** An early-stopping module allows training to be finished prematurely after the loss reduction falls below a certain threshold for subsequent boosting rounds.

- **Improved split algorithm.** An approximate and sparsity-aware algorithm for finding the best split in each node (see Equation 2.8) outperforms the greedy algorithm described in Subsection 2.2.2 in terms of computation time.

- **Learning rate.** The introduction of a learning rate which shrinks the contribution of each tree heavily increases the accuracy (Hastie et al., 2009). This idea was already introduced by Friedman (2001).

Feature importance of gradient boosted trees

The feature importance for gradient boosted trees is calculated in the same way as for random forests (see Equation 2.20)

$$\mathcal{T}_m^{BT} = \frac{1}{B} \sum_{b=1}^{B} \mathcal{T}_m(f_b^{RT}) \tag{2.28}$$

In contrast to random forests, the variable-split selection for boosted trees does not include a single feature by chance, thus leading to more sparse feature selection (Hastie et al., 2009). While a random forest would rank all features according to their importance and allow several correlated features to be ranked high, a boosted tree model would ignore correlated features in the importance after selecting the most powerful ones.

Advantages and disadvantages of gradient boosted trees/XGBoost

Advantages Again, XGBoost (and a gradient boosted tree model in general) inherits many advantages from the design of decision trees, which were already discussed in Subsection 2.2.2. However, they also have some more benefits.

- **Accuracy.** In comparison to the other methods, the performance of gradient boosted trees is rated best. However, the deviation from a random forest is quite small (Hastie et al., 2009).
- **Customizability.** The model complexity can be controlled through many hyperparameters, and regularization is easily applicable (Chen and Guestrin, 2016; Johnson and Zhang, 2013).
- **Performance.** Although different trees cannot be grown in parallel, XGBoost achieves a major speedup by parallelizing within a single tree achieving lower training times than comparable implementations (Chen and Guestrin, 2016).

Disadvantages In contrast, some advantages of decision trees are lost in a XGBoost model.

- **Complex to understand.** An XGBoost model has many hyperparameters that have to be tuned and usually are optimized through cross-validation. For beginners, it might not be clear how each parameter influences the outcome of the model.
- **Interpretability.** Of all models, XGBoost is the least interpretable model. That is because single trees are fit to the residuals, and thus inspection of them has no valid meaning anymore.

2.2.5 Model evaluation

After understanding how these algorithms work, it is nevertheless still important to evaluate the output of such models. Thus, in this section, we introduce common model-agnostic methods to interpret models and their predictions.

Confusion matrix

The confusion matrix and metrics which can be derived from it are frequently used to evaluate a model (Hoens and Chawla, 2013). Table 2.1 shows an exemplary confusion matrix for binary classification.

Table 2.1: Exemplary binary confusion matrix. TP are true positives, FN are false negatives, FP are false positives and TN are true negatives.

		Predicted label $\hat{y}$		
		1	0	Total
Observed label y	1	TP	FN	TP+FN
	0	FP	TN	FP+TN
	Total	TP+FP	FN+TN	TP+FP+FN+TN

Many different metrics can be computed from the confusion matrix, all looking at different aspects and having different benefits and drawbacks. Murphy (1996) discussed how crucial it is to choose a suitable metric for the verification of forecasts. Naively, a typical metric one would look at is the

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}.$$

However, especially for imbalanced datasets, *Accuracy* is not a suitable metric since it puts too much emphasis on the major class (Hoens and Chawla, 2013). A model that solely predicts $\hat{y} = 0$ achieves an *Accuracy* of 99% if each class-1 observation faces a hundred class-0 ones. Thus, for imbalanced datasets, one would instead consider the

$$Precision = \frac{TP}{TP + FP},$$

which measures how many of all positively predicted cases were correct. Second, the

$$Recall = \frac{TP}{TP + FN},$$

measures how many of all class-1 observations were correctly predicted. Finally, the F1-score combines both metrics into one

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall},$$

and is nothing but the harmonic mean between *Precision* and *Recall*. Here, we followed the convention and defined class-1 to be the rare class.

Depending on the problem, a different metric might be of interest (e.g., in pandemics, higher *Recall* might be preferable since fewer false negatives will be made). Since all metrics depend on the threshold c, one usually optimizes c for the specific metric of interest. Also note, that each metric might be known under a different name in different fields. In meteorology, one would translate *Precision* to the Correct-Alarm-Ratio (CAR) and *Recall* to the Probability-Of-Detection (POD).

Another popular measure is the so-called Receiver-Operator-Characteristic curve (ROC curve) because it aggregates the confusion matrix for all thresholds. A ROC curve can be constructed with the false positive rate

$$FPR = \frac{FP}{FP + TN} \tag{2.29}$$

on the x-axis and the true positive rate (an alias for *Recall*)

$$TPR = \frac{TP}{TP + FN} \tag{2.30}$$

on the y-axis of a diagram (Hoens and Chawla, 2013). One point $(FPR(c), TPR(c))$ is added to the diagram for an uniformly distributed set of thresholds $c \in [0, 1]$. Several potential ROC curves are shown in Figure 2.5. Thus, with a single plot one can evaluate a model for all potential thresholds c.

Another widely used metric, which in contrast to the metrics above is threshold independent, is the so-called ROC Area Under Curve (AUC). The AUC integrates the area under the ROC curve. A perfect model would follow a curve which which passes through the points $(0, 0)$, $(0, 1)$ and $(1, 1)$ and thus score $AUC = 1$. A model which just guesses randomly will passes only through the points $(0, 0)$ and $(1, 1)$ and consequently score $AUC = 0.5$.

Again, since the FPR includes true negatives, Davis and Goadrich (2006) have argued that ROC curves (and the AUC correspondingly) present an overly optimistic view on the model's performance in the case of large class imbalance. Instead, they suggested to rely on *Precision*, *Recall*, and the corresponding *Precision-Recall* curve. In our case, the interest in the ROC and AUC stemmed from the ability to compare our model to other scientists work.

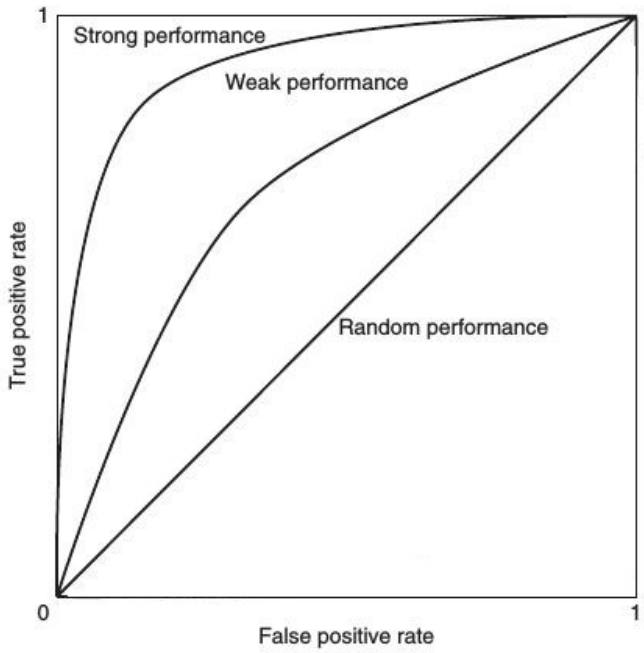


Figure 2.5: Shown are several exemplary ROC curves. The closer a curve gets to $(0,1)$ the better the model. The AUC can be determined by integrating over the curve. Figure taken from Hoens and Chawla (2013).

Chapter 3

Data

3.1 Future development of foehn winds

In the next sections, we describe the data which was employed during our analysis. In Subsection 3.1.1, we look at the ERA-Interim reanalysis data. In Subsection 3.1.2, the structure of the CESM simulation data is described. In Subsection 3.1.3, the preprocessing of the observational foehn time series will be explained.

3.1.1 ERA-Interim data

The ERA-Interim (ERAI) reanalysis (Dee et al., 2011) was produced by combining observational data of the past decades with an NWP model by means of data assimilation. ERAI describes the state of the atmosphere with a spatial horizontal grid resolution of one-degree (approximately 80 km in our area of interest) on 60 vertical levels and a temporal resolution of six hours (i.e., at 0:00, 6:00, 12:00, and 18:00). All features which were used for the training and testing are derived from the ERAI reanalysis. In our case, we used data from January 1981 until the end of March 2019.

Since, for our analysis, the synoptic conditions in the Alpine region were relevant, we decided to limit the area of interest to a number of grid points over Europe. Those grid points would lie in the interval $[42°, 50°] \times [0°, 15°]$ latitude and longitude, respectively. This choice is also motivated by the relevant variables in the Widmer test (Widmer, 1966; Courvoisier and Gutermann, 1971). However, since we would make predictions on the CESM data, we interpolated the variables from the ERAI grid to the CESM grid (see Figure 3.1). We interpolated this way (and not from CESM to ERAI) due to the more coarse CESM grid resolution. In total, this procedure left us with 104 horizontal grid points per variable.

Due to the enormous amount of variables in ERAI, we needed to limit our-

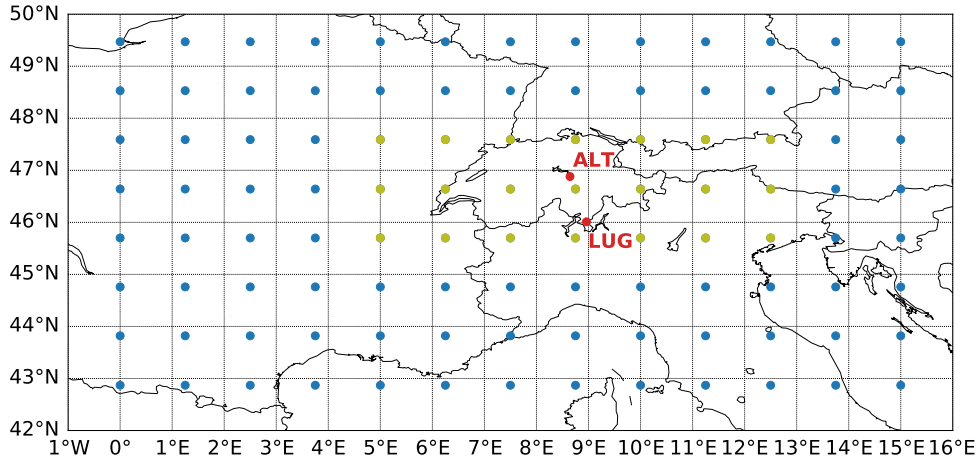


Figure 3.1: Grid Points of the CESM Grid (blue). While the ERAI grid is defined on full degrees (intersection of the lines), CESM is slightly more coarse and shifted. The locations of Altdorf (ALT) and Lugano (LUG) are marked in red. The olive points indicate the grid points in the Alps which we removed to achieve better generalization to CESM (see Section 4.1 for details).

Table 3.1: Selected raw variables and their corresponding chosen pressure levels.

Variable	Description	Pressure levels [hPa]
SLP	Sea level pressure	sea level
Z	Geopotential height	850, 700
T	Temperature	900, 850, 700
U	Zonal wind component	700, 500
V	Meridional wind component	700, 500

selves to the essential. Thus, we used a physically-motivated approach and selected variables that are known to be relevant for foehn. Sprenger et al. (2017), Richner and Hächler (2013), Zweifel et al. (2016) and Gerstgrasser (2017) defined sea level pressure differences ΔSLP, geopotential height differences ΔZ, potential temperature differences $\Delta\theta$ and wind flow $U\&V$ to be important. The raw variables read from the ERAI data are described in Table 3.1. We decided to look at variables at sea level, 900 hPa, 850 hPa, 700 hPa and 500 hPa. Here, 900 hPa resembles a layer close to the surface, 850 hPa resembles a layer about halfway to the Alpine crest, 700 hPa is at the Alpine crest layer, and 500 hPa is a layer high above the Alpine crest. A topography plot of ERAI is shown in Figure 3.2. Note how the Alps are represented as an obstacle smaller than in reality. However, the synoptic-scale conditions which cause foehn (i.e., pressure systems) are unlikely to be affected by this.

The calculation of the final variables was straightforward and can be found in Appendix A. Finally, we were left with the variables shown in Table 3.2. Each row in the final dataset represented one observation (i.e., state of the

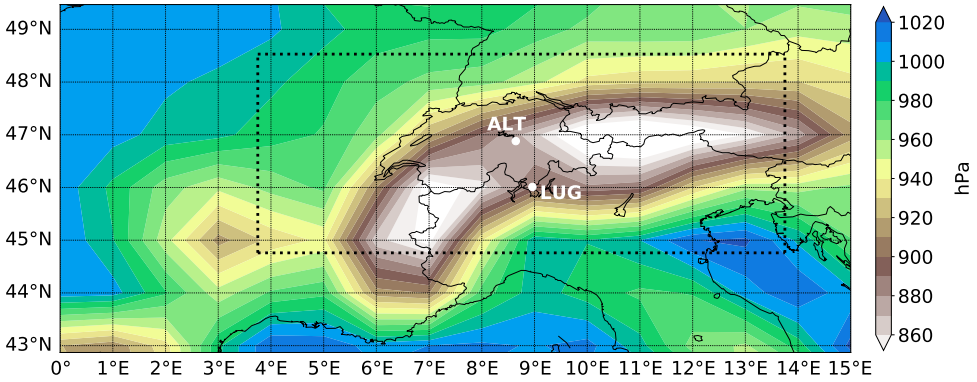


Figure 3.2: Topography of ERAI in terms of an averaged surface pressure. The locations of the south foehn station Altdorf (ALT) and the north foehn station Lugano (LUG) are shown in white. The black dotted frame indicates the area within which we would remove the features for the CESM prediction (see Section 4.1 for details).

Table 3.2: The final features. The raw SLP, T and Z values got removed, after the more important variable differences got calculated.

Variable	Description	Pressure levels [hPa]	# features
ΔSLP	Sea level pressure diff.	sea level	5356
ΔZ	Geopotential height diff.	850, 700	10712
$\Delta\theta_{hor}$	Hor. pot. temperature diff.	850	5356
$\Delta\theta_{ver}$	Ver. pot. temperature diff.	850, 700	208
U	Zonal wind comp.	700, 500	208
V	Meridional wind comp.	700, 500	208

atmosphere) at a given date-time (e.g., 29/11/1985 6:00). Each column in the final dataset represented one feature at a specific grid point (or difference) and pressure level (e.g., V at 48.53°N and 10°E on 700 hPa).

3.1.2 CESM data

In contrast to the ERAI reanalysis, the CESM model (Kay et al., 2015) does not rely on past observational data and is a freely running climate simulation with a slightly different horizontal grid (see Figure 3.1) on 30 vertical levels and six-hour temporal resolution. The CESM simulation was applied to two scenarios. First, for a present-day climate (CESM-p), keeping parameters like CO_2 concentration fixed at a present level. Second, for a future climate (CESM-f), external forcing is adjusted to the expected future levels. Here, we relied on the RCP8.5 scenario discussed in Van Vuuren et al. (2011). Each simulation (CESM-p and CESM-f) consisted out of 35 ensemble members, each of them one possible realization of the weather conditions during the ten-year period from 1991 until 2000 for present-day climate and 2091 until 2100 for future climate. Here, each ensemble member was initialized

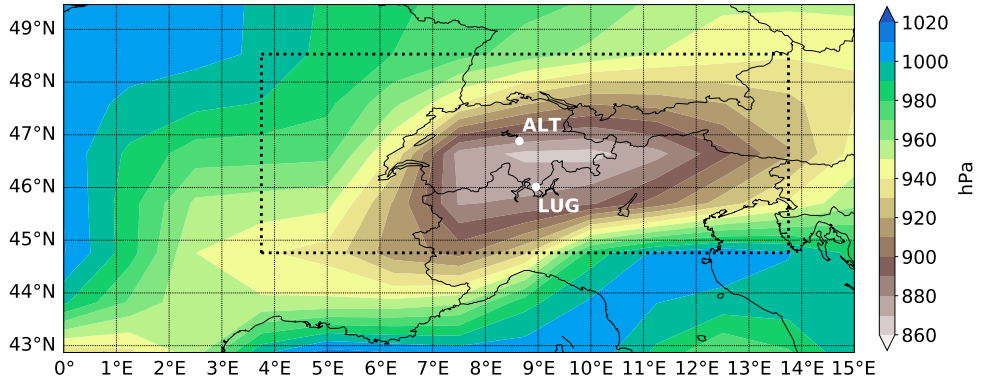


Figure 3.3: Topography of CESM in terms of an averaged surface pressure. The locations of the south foehn station Altdorf (ALT) and the north foehn station Lugano (LUG) are shown in white. The black dotted frame indicates the area within which we would remove the features for the CESM prediction (see Section 4.1 for details).

with slightly different initial conditions and then integrated forward. Thus, even though the ensemble members of a given period represent the same climate, the exact weather conditions will differ. For more details regarding the implementation, see Röthlisberger et al. (2020).

The CESM model data for both scenarios was retrieved similarly to the ERAI data. Apart from the different grid (see Figure 3.1), the CESM data was not already available at pressure levels. Thus, we used the pressure values at the 30 model levels to linearly interpolate the raw variables given in Table 3.1 to their corresponding pressure levels. In case interpolation was not possible (i.e., if the pressure level was below the lowest model level), we decided against extrapolating this variable for this pressure level and grid point. That is because, in this case, the pressure level is very likely to intersect with the topography.

A topography plot for CESM can be found in Figure 3.3. Note how the Alps are more coarsely resolved compared to ERAI. In the end, the derived variables were calculated in the same way as for ERAI (see Appendix A).

3.1.3 Foehn data

For the training, we needed a label for the foehn target variable $y^{(i)}$. Since we dealt with a binary classification problem, we needed to associate each ERAI observation (i.e., each state of the atmosphere at 0:00, 6:00, 12:00, and 18:00) with its target variable

$$y^{(i)} = \begin{cases} 1 & \text{if foehn is prevalent} \\ 0 & \text{else.} \end{cases} \tag{3.1}$$

The foehn time series was gratefully made available by the Swiss national weather service (MeteoSwiss; www.meteoswiss.ch). For south foehn, the time series contained the Duerr-Index (Duerr, 2008) at Altdorf from January 1981 until December 2019 in ten-minute intervals. Duerr (2008) separates between foehn, mixed-foehn and non-foehn cases.

Since we were interested in the synoptic situation of foehn, we treated mixed-foehn as normal foehn. For mixed-foehn cases, foehn is eventually prevalent. However, local conditions like not completely eroded cold pools prevent the foehn from completely breaking through (Duerr, 2008). Due to the coarse grid resolution, these local conditions are not captured in ERAI or CESM.

In the next step, we used the values in the ten-minute time series to create the label $y^{(i)}$ for each ERAI time step (i.e., at 0:00, 6:00, 12:00, and 18:00). In Gutermann et al. (2012) it has been shown that the best transition from ten-minute intervals to hourly observations in the morning, noon, and evening can be achieved if at least 4 out of 6 intervals indicate foehn. Moreover, it is argued that this 4-out-of-6 rule is invariant under a 30-minute shift to the past or future. Since we possibly wanted to look at the situation amid foehn, we defined that foehn is prevalent at 0:00, 6:00, 12:00, and 18:00 if at least 40 out of 60 minutes, which are centered symmetrically around the given time, showed the occurrence of a foehn wind.

The mean foehn duration for all foehn events in the dataset was approximately 5.5 hours. That coincided with Orlanski (1975), who found that synoptic situations (e.g., pressure systems or fronts) are usually persistent from hours to days. Thus, even though we dealt with only four measurements per day, we had a good chance of capturing a foehn event.

For north foehn, we followed the same approach as for south foehn, only here we had data available from January 1983 until December 2019. The data was generated by Cetti et al. (2015), who followed the same methodology as Duerr (2008).

3.2 Impact of foehn on forest fires

3.2.1 Forest fire data

The forest fire data was gratefully made available by the Swiss Institute for forest, snow, and landscape (WSL, www.wsl.ch). For details regarding the database, see Pezzatti et al. (2010). The database included all forest fires in Switzerland from January 1980 until December 2019. In total, 4599 fires were recorded during this period. The spatial distribution of fires can be seen in Figure 2.2.

Each fire in the dataset was further specified by 36 characteristics. Amongst the most important ones were the burned area, the start and end timestamp, the accuracy of those timestamps, the coordinates, municipality, and the most probable cause. Since many values showed themselves to be missing, we first had to cleanse the data and hence make some assumptions described in the next paragraph.

First, we decided to keep only fire data where we knew the start and end timestamp of a fire at least to an accuracy of an hour. Second, in collaboration with the experts from WSL, we learned that entries with missing values in the burned area resembled small fires. Thus, we imputed the missing burned area fields with a value of 0.01 ha. Lastly, where the exact coordinates of the fire were missing, we imputed the values with the center coordinates of the municipality. Here, we retrieved the coordinates automatically via the Nominatim API and performed a coordinate transformation from WGS84 coordinates to the Swiss LV03 coordinates (see Swisstopo (2016)).

3.2.2 Foehn data

In total, foehn data from 39 meteorological stations was available to us in form of the Duerr index (Duerr, 2008) at ten-minute resolution (see Subsection 3.1.3). Furthermore, measurement data was available for the wind speed FF, wind gust peak FFX, and temperature TT, which we used to assess foehn strength. The location of all foehn observation stations is shown in Figure 2.2.

Again, for each station, we preprocessed the data, e.g., cleansed missing values. In contrast to the above section about foehn development (Subsection 3.1.3), we treated mixed-foehn as non-foehn cases. The reason for this was that now we were concerned with the mesoscale conditions of foehn. For mixed-foehn cases, the probable existence of a cold pool in the valley prevents the foehn from breaking through and showing its typical characteristics (see Figure 1.1).

Chapter 4

Implementation

4.1 Future development of foehn winds

We divided this chapter and the corresponding ones in the Results and Discussion (see Chapter 5) into multiple parts. Note how each part corresponds closely to an objective described in Subsection 1.1.3. First, we look at the most important features to identify south foehn from coarse NWP data. Then, we describe the set-up and evaluation of a statistical model on ERAI data. In the next step, we focus on generalizing the model's predictions to CESM data. In the fourth part, we compare the predictions on all datasets, making sense of a potential future development of south foehn. Lastly, we discuss the same for the analysis of north foehn.

Synoptic-scale foehn conditions

To identify the most relevant features for foehn, we investigated feature importances on the variables listed in Table 3.2. For ΔSLP, ΔZ, $\Delta\theta_{hor}$, $\Delta\theta_{ver}$, and $U\&V$, we separated the variables into five different datasets, respectively. Now we trained a random forest with $B = 1000$ and $d_{max} = 7$ on each dataset and selected the 30 most important variables according to the feature importance (see Equation 2.20). The advantage of this method lies in the fact that the selected variables (e.g., SLP differences) can directly be compared to the theory. In order to do so, we plotted the selected features over a composite plot of the mean conditions prevalent during foehn (i.e., where $y^{(i)} = 1$ as defined in Equation 3.1).

Foehn predictability on ERAI

First, we split the ERAI data $\boldsymbol{D}_{ERA}$ into training set $\boldsymbol{D}_{train}$ and test set $\boldsymbol{D}_{test}$. We decided to use the years from 1991 until 2000 for $\boldsymbol{D}_{test}$ since CESM-p simulates the same period. Consequently, the remainder of the data, which

originated from the years 1980 until 1990 and 2001 until 2019, was used for $\boldsymbol{D}_{train}$.

Next, we trained a range of machine learning algorithms on $\boldsymbol{D}_{train}$ and evaluated their performance on $\boldsymbol{D}_{test}$. In detail, we utilized the following models: Random Forests, Gradient Boosted Trees (i.e., XGBoost, LightGBM, CatBoost), Deep Neural Networks (DNNs), K-Nearest Neighbors (KNN) with Principal Component Analysis (PCA) preprocessing, and an Elastic Net. Furthermore, due to the spatial nature of the data, we also employed a Convolutional Neural Network (CNNs) to the data. The hyperparameter optimization was done using a three-fold cross-validation on $\boldsymbol{D}_{train}$. We optimized over the various individual hyperparameters of the algorithms (such as depths, learning rates, or regularization parameters). The best model was determined using the summed logistic loss for all samples (compare Equation 2.24).

In the end, the best performance was obtained by the XGBoost algorithm, even though the results of some other algorithms were not substantially worse. Primarily the CNN approach also worked considerably well. Nevertheless, the benefits of XGBoost models over CNNs originate from less resource-intense training procedures and more straightforward interpretability due to the intrinsic feature importance (see Equation 2.28). Finally, the results of XGBoost on the whole dataset would serve as our baseline, against which we could compare the following steps.

Generalization to CESM

After removing the features, which were not available in CESM (due to a likely intersection with the topography; see Subsection 3.1.2), we used a straightforward approach and applied a trained XGBoost model to CESM data. However, as it turned out, this approach worked only in very limited ways since we generally predicted very little foehn occurrence on CESM-p. We wondered whether this could stem from overfitting to ERAI data and hence having poor generalization to CESM data. For this reason, we employed a range of additional preprocessing measures.

Quantile rescaling To compensate for any bias or deviation in the representation of features between ERAI and CESM-p (induced by different model topography and necessary grid interpolations), we separately rescaled both datasets. Here, we chose a non-parametric approach and transformed a feature to its quantile representation. That would not affect the results within ERAI since the quantile function $q : \mathbb{R} \to [0,1]$ is strictly monotone, and trees are invariant under such a transformation. Moreover, to avoid model leakage, the samples in $\boldsymbol{D}_{test}$ were rescaled with the q, which was fitted and applied to $\boldsymbol{D}_{train}$. For CESM-f, we must rescale each feature with the func-

tion q we obtained from the corresponding feature in CESM-p because we wanted to investigate a shift in the feature distribution between CESM-p and CESM-f. Hence, if we had rescaled CESM-f with its own $\tilde{q}$, we would have eliminated part of what we liked to measure. Then, CESM-p and CESM-f would have a very similar feature distribution.

Disregarding local features We wanted our model to classify the synoptic-scale weather situation. Thus, we decided to exclude features in the Alpine region, which are likely to be represented differently due to the deviating topography between ERAI and CESM. In particular, we excluded the 21 grid points inside the rectangle between [44.76°N, 3.75°E] and [48.53°N, 13.75°E] from $\boldsymbol{D}_{ERA}$, $\boldsymbol{D}_{CESMp}$, and $\boldsymbol{D}_{CESMf}$ (see Figure 3.1, Figure 3.2, and Figure 3.3).

Monthly models We moved from employing one XGBoost model for the whole year to utilizing twelve XGBoost models, one for each month, respectively. This allowed us to adjust the threshold so that *Precision* and *Recall* would be balanced (i.e., false positives and false negatives occurred with the same frequency). Furthermore, this also allowed for inspection of the individual models for potentially varying feature importances over the year.

Constrained optimization of the loss function Lastly, as the potentially most impactful step, we adjusted the loss function l for the XGBoost models. For ERAI samples, we continued to use the standard logistic loss from Equation 2.24

$$l_{LL}(y^{(i)}, p^{(i)}) = -(y^{(i)} \log(p^{(i)}) + (1 - y^{(i)}) \log(1 - p^{(i)})).$$

Now, we introduced an additional term, which would be optimized on the CESM-p samples during the training procedure. Since we did not have any labels available on CESM directly, we formulated this term as

$$l_{SME} = \lambda \left(\frac{\sum_{\boldsymbol{D}_{CESMp}} p^{(i)}}{\sum_{\boldsymbol{D}_{CESMp}}} - \frac{\sum_{\boldsymbol{D}_{train}} y^{(i)}}{\sum_{\boldsymbol{D}_{train}}} \right)^2 = \lambda \left(\mu_{CESMp} - \mu_{train} \right)^2, \tag{4.1}$$

where $p^{(i)} = 1/(1 + \exp(-z^{(i)})) = 1/(1 + \exp(-f_{b-1}^{BT}(\boldsymbol{x}^{(i)})))$ is again the prediction for sample i after applying the logistic function (Equation 2.22). We will call this term *Squared Mean Error* from here. This term acted as a form of constraint to the optimization procedure (similarly as regularization) to make the mean of $\boldsymbol{D}_{CESMp}$ as close to the mean on $\boldsymbol{D}_{train}$ for each month. With the parameter λ, we could control how strong we wanted to enforce this constraint. For $\lambda = 0$, we would completely disregard the constraint. Thus, the XGBoost models were forced into a trade-off. On the one

hand, they were forced to select features which allowed for an accurate prediction on $\boldsymbol{D}_{train}$ through Equation 2.24. On the other hand, the models were penalized for selecting features which made the predicted mean on $\boldsymbol{D}_{CESMp}$ too unequal compared to the observed mean on $\boldsymbol{D}_{train}$ through Equation 4.1. Hence, we manged to reduce overfitting on $\boldsymbol{D}_{train}$ and enforced better generalization to $\boldsymbol{D}_{CESMp}$.

For the training procedure, we required the gradient and trace of the hessian of Equation 4.1 with respect to the prediction of the regression trees $z^{(i)}$ (see Equation 2.26). After some calculation, the gradient was determined as

$$\frac{\partial l_{SME}}{\partial z^{(i)}} = \frac{2\lambda}{N_{CESMp}} \underbrace{p^{(i)}\left(1 - p^{(i)}\right)}_{(*)} \underbrace{\left(\mu_{CESMp} - \mu_{train}\right)}_{(**)}. \tag{4.2}$$

The trace of the hessian can be written as

$$\begin{aligned}\frac{\partial^2 l_{SME}}{(\partial z^{(i)})^2} = \frac{2\lambda}{N_{CESMp}} p^{(i)}\left(1 - p^{(i)}\right) \Big[&\left(1 - p^{(i)}\right)\left(\mu_{CESMp} - \mu_{train}\right) \\ &+ p^{(i)}\left(\mu_{train} - \mu_{CESMp}\right) \\ &+ \frac{1}{N_{CESMp}} p^{(i)}\left(1 - p^{(i)}\right)\Big].\end{aligned} \tag{4.3}$$

The different factors in the gradient and hessian allowed for deeper insight into what we optimized. On the one hand, we find that $(*) \to 0$ for $p^{(i)} \to \{0, 1\}$. Consequently, the gradient vanishes for samples where the algorithm is certain about its prediction. On the other hand, we find that $(**) \to 0$ for $\mu_{CESMp} \to \mu_{train}$. Thus, the gradient also disappears if we approximate the observed mean on $\boldsymbol{D}_{CESMp}$. Furthermore, $(**)$ flips the sign of the gradient depending on whether we over- or underpredict foehn on $\boldsymbol{D}_{CESMp}$. A similar interpretation can be found for the second-order correction terms in the hessian. Here, the first term makes the model correct its prediction for samples with $p^{(i)}$ slightly smaller than 0.5. The second term affects samples with $p^{(i)}$ slightly larger than 0.5. The direction of the second-order corrections are again determined by the sign of $\mu_{CESMp} - \mu_{train}$. The third term is negligible in our case due to its small magnitude $1/N_{CESMp} \approx 10^{-4}$.

Finally, we employed the XGBoost models on CESM data. First, we selected the most important 250 features from an XGBoost model trained on all features of $\boldsymbol{D}_{train}$ via three-fold cross-validation. Then, we used those features to retrain the monthly models. Here, the monthly models had $B = 200$ and $d_{max} = 4$. To utilize also information from other months and having already some predictive skill on CESM (often called a "warm-start"), the first

ten trees were trained on the yearly $\boldsymbol{D}_{train}$ without the constraint condition. In contrast, the next 190 trees are trained only on monthly data under the constraint condition with $\lambda = 60000$. Afterward, we employed the models to make predictions for all 35 ensemble members of CESM-p and CESM-f. Then, we generated the same composite plot for CESM-p and CESM-f as for ERAI. However, this time, we used the predicted label to determine a foehn situation (i.e., where $\hat{y}^{(i)} = 1$). The feature importances were generated by summing the 30 highest ranked features importances from each monthly model.

Comparison of monthly foehn frequencies

Finally, we compared the monthly aggregated data for the different datasets. To do so in a structured approach, we contrasted the data stage-wise for their statistical significance:

1. **Observational data vs. ERAI predictions.** Since in this case, we knew the correct label at each observation, we used a forecast verification approach. Here, we relied on the confusion matrix and derived metrics (see Subsection 2.2.5).

2. **ERAI predictions vs. CESM-p predictions.** Due to the fact that CESM-p contained 350 years of data, ERAI however only 10, this transition was more challenging. Nevertheless, so we argue, can ERAI be viewed as one potential ensemble member of CESM-p, i.e., one potential realization of the weather. Thus, we randomly sampled a month from each of the ten years in the CESM-p data and calculated the mean foehn frequency. Then, we repeated this procedure 1000 times and compared the resulting Gaussian distribution with the mean foehn frequency of ERAI. Of course, due to the constraint optimization, the ERAI frequency should, by design, be close to the CESM-p frequency. However, the chance existed that the XGBoost models failed to find meaningful features that would generalize well. If the ERAI value fell outside the second standard deviation of the distribution, we rejected the hypothesis that they follow the same underlying distribution. In this case, a meaningful transition from ERAI to CESM-p was not possible.

3. **CESM-p predictions vs. CESM-f predictions.** In the last test, we dealt with the same amount of data points. Thus, we used a Wilcoxon rank-sum test and calculated a p-value for each month to judge whether the distributions differ significantly. As global rate of error we used $\alpha = 0.05$. Since we dealt with a multiple testing scenario, we further applied the Bonferroni correction, thus reducing the effective rejection level for each month to $\tilde{\alpha} = 0.0042$. If the p-value fell below $\tilde{\alpha}$, we

rejected the null-hypothesis that the two samples follow the same underlying distribution and conducted statistical significance.

North foehn development

For the north foehn analysis, we could fully utilize the beauty of our approach. The ERAI, CESM-p, and CESM-f data stayed utterly untouched. The only thing we needed to change were the target labels $y^{(i)}$. We then applied the same steps in the same order as for the south foehn analysis. Here, we selected the 250 most important features for the training of the monthly XGBoost models, and used the parameters $B = 200$, $d_{max} = 4$, and $\lambda = 90000$.

4.2 Impact of foehn on forest fires

The implementation for this work package mainly consisted of the data preprocessing, which is extensively described in Section 3.2. After both data sets were prepared, we mapped each fire to the spatially closest meteorological station. If the next station was further away than 20 km, we refused to associate this fire with any station. Visually, this procedure is depicted in Figure 4.1. That left us with about 3300 fires, which can be evaluated regarding their foehn activity.

Note, however, that even if a fire could be mapped to a meteorological station spatially, this did not imply that we could associate the fire with foehn temporally. Most stations started to record foehn relevant parameters several years after 1980 and showed missing values eventually. If any foehn values were missing during the fire duration or the period before (see below), this fire would not count into the statistic. On average, this left us with around 2300 fires, depending on the objective.

Following the objective questions from Subsection 1.2.3, we investigated two different time periods.

Prior fire ignition

We calculated the accumulated foehn minutes 24 and 48 hours prior to ignition. Then, we counted how often fires are associated with a particular duration of foehn and binned the results by foehn minutes. Further, we normalized each bin with the overall occurrence of the bin-specific foehn minutes due to the rarer prevalence of longer foehn events. Here, this occurrence was calculated with a sliding window operation for each measurement station and then averaged over all stations. For readability, the axis was then rescaled with the minimum of the resulting values.

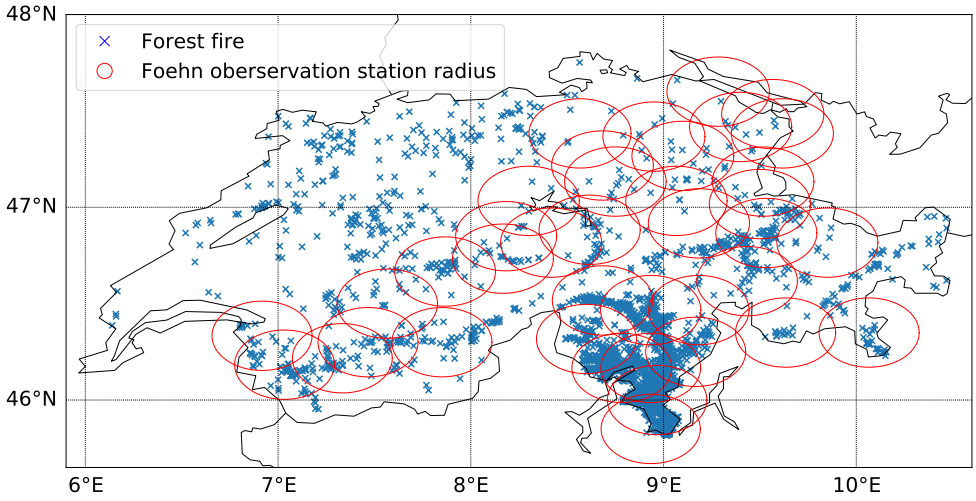


Figure 4.1: Each forest fire from 1980 until 2019 (blue) within a 20 km radius around a meteorological station (red) was mapped to the closest station. Note how this distance is distorted to an ellipse due to the WGS84 projection. The actual distance calculation was performed in LV03 coordinates, however, due to readability this plot is depicted in WGS84 coordinates. See Swisstopo (2016) for details.

Afterward, we repeated the same procedure for foehn temperature increase ΔT_{foehn}. We counted how often fires are associated with specific binned foehn temperature increases, again normalizing with the general occurrence of such a temperature increase. Here, ΔT_{foehn} is defined as the mean over all temperature values in the last 24 or 48 hours which showed foehn activity minus the mean over all temperature values in the last 24 or 48 hours which did not show foehn activity

$$\Delta T_{foehn} = \mu_{foehn}(T) - \mu_{non-foehn}(T) = \frac{\sum_{y^{(i)}=1} T^{(i)}}{\sum_{y^{(i)}}} - \frac{\sum_{y^{(i)}=0} T^{(i)}}{\sum_{y^{(i)}}}. \tag{4.4}$$

Post fire ignition

After the start of a fire, we calculated the accumulated foehn minutes for the next six and twelve hours. Here, it is vital to watch out for a potential confounder/mediator fallacy shown in the causal diagram in Figure 4.2.

If a fire is shorter than six hours, it would make sense to only count the foehn minutes during this period. However, in this case, we would introduce an edge in our graph from *Fire duration* to *Foehn duration*, since *Fire duration* limits the maximum minutes of *Foehn duration*. That introduces a confounder bias to the causal effect from *Foehn duration* to *Burned area*: A short fire duration limits the foehn minutes, while at the same time, it is likely associated with a small burned area. Usually, one would now control for *Fire duration* to eliminate this bias.

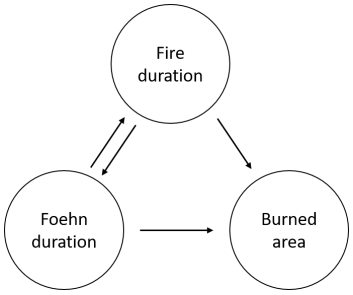


Figure 4.2: Causal diagram showing the relation between the *Foehn duration* and the *Burned area* for the first six or twelve hours after ignition, when capping the maximum *Foehn duration* at the *Fire duration*.

Unfortunately, however, in our case *Fire duration* might also be a mediator between *Foehn duration* and *Burned area*: More foehn minutes during the ignition period lead to a longer fire duration, which in turn increases the burned area. In this case, we cannot control for the mediator since it potentially eliminates part of the effect we want to measure.

Our resolution for this was to always consider the full six (twelve) hours after ignition. For example, a fire extinguished after three hours could still have a value of six (twelve) hours for *Foehn duration*. This way, *Fire duration* no longer limits *Foehn duration*, and we can delete this edge in our graph. However, this also leads to underestimating the total effect, since the foehn minutes after fire extinguishment no longer influence the fire causally, yet are attributed to a long *Foehn duration* bin. For this reason, the results depict a lower bound.

We evaluated the effect from *Foehn duration* on *Burned area* by visualizing the results in the form of a boxplot. Furthermore, we disaggregated the data and looked at them from the viewpoint of different fire regimes, foehn locations, and decades. Lastly, we performed a statistical Wilcoxon rank-sum test to evaluate significance in the different distributions, i.e., between fires without foehn occurrence (non-foehn fires) and fires with foehn occurrence (foehn fires). Here, we used a global rate of error $\alpha = 0.05$. Whenever we faced a multiple testing scenario, we applied the Bonferroni correction.

Again, we repeated the same analysis for the mean foehn strength, i.e., wind speed FF and gusty wind peaks FFX, during the first six or twelve hours of a fire as measured by the closest meteorological station. Here, we took only the foehn fires into account. FF was defined by the mean of all wind measurement values, which showed foehn occurrence

$$FF = \mu_{foehn}(FF) = \frac{\sum_{y^{(i)}=1} FF^{(i)}}{\sum_{y^{(i)}}}. \tag{4.5}$$

FFX is defined analogously.

Chapter 5

Results & Discussion

5.1 Future development of foehn winds

Recall, that every part in this chapter is aimed at answering one particular question raised in the objectives (see Subsection 1.1.3).

5.1.1 Results

Synoptic-scale foehn conditions

As stated in Section 4.1, we identified the most important features on ERAI by training a random forest on each variable type. This allowed us to obtain first insight into the data. Figure 5.1 shows the 30 most important sea level pressure differences, which were plotted over a composite plot of all observed foehn events (i.e., where $y^{(i)} = 1$). We also performed the same analysis for ΔZ (see Figure 5.2), $\Delta\theta_{hor}$ (see Figure 5.3), $\Delta\theta_{ver}$ (see Figure 5.4), and the wind components U and V (see Figure 5.5). Lastly, we also investigated the composite plot for predicted foehn cases (i.e., where $\hat{y}^{(i)} = 1$), however, almost no difference was discernible.

Although all shown features are amongst the features which have the highest predictive power on ERAI, these plots have to be taken with a grain of salt. First, the shown important features do not imply that foehn cannot be identified from other features that are not shown. However, using these other features will come at the cost of more miss-classified samples. Second, due to the high spatial correlation of the features, one cannot identify "the" most crucial feature. If excluding this feature from the training, the algorithm will select another spatially close feature at little cost of accuracy. Furthermore, the spatial correlation also implies that after having used a feature during training, other spatially close features will likely contain little additional information. Consequently, it is vital to view the plots as a

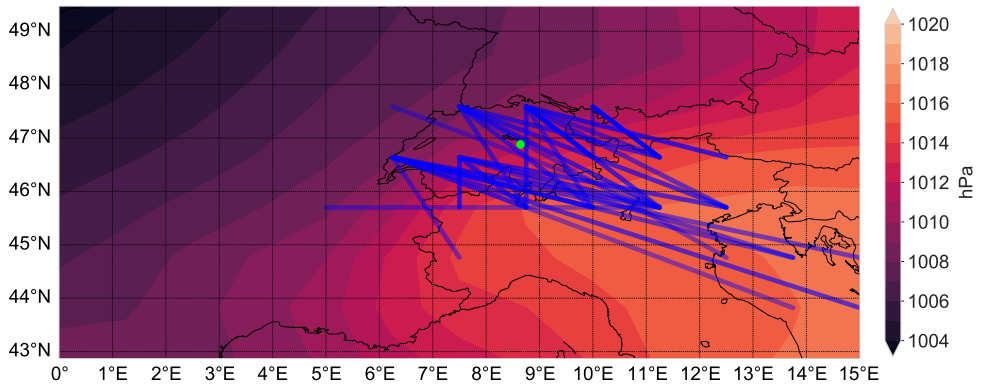


Figure 5.1: Shown are the 30 most important sea level pressure differences ΔSLP over the mean SLP weather condition in ERAI for all observed foehn events. The more opaque the blue line, the more important is the specific difference according to the random forest feature importance. The green dot marks the location of Altdorf.

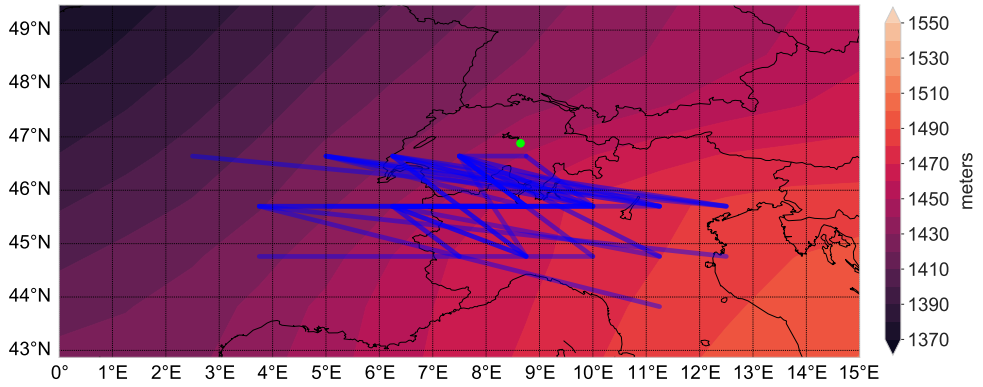


Figure 5.2: Same as Figure 5.1 but for ΔZ on 850 hPa.

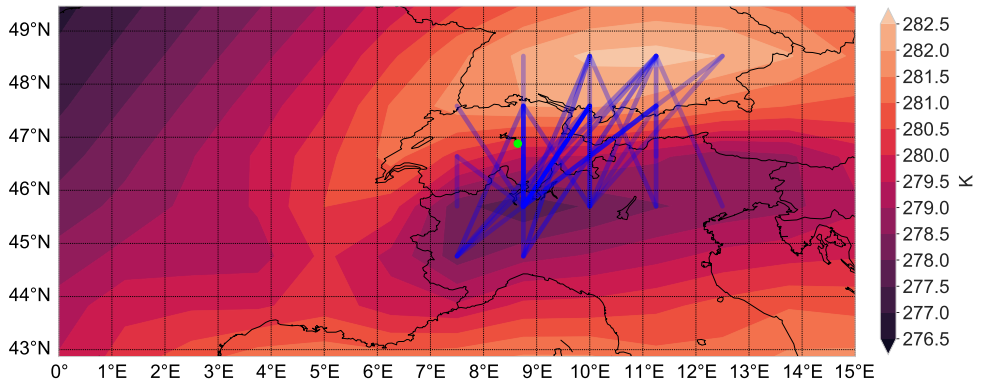


Figure 5.3: Same as Figure 5.1 but for $\Delta\theta_{hor}$ on 850 hPa.

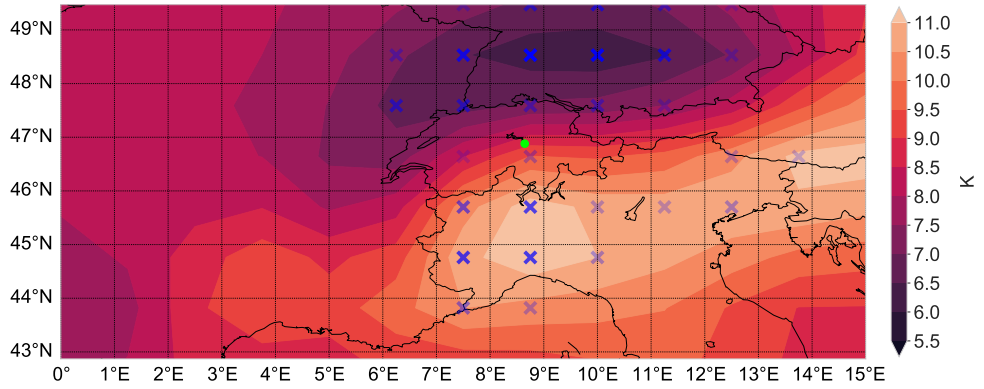


Figure 5.4: Same as Figure 5.1 but for $\Delta\theta_{ver}$ between 900 and 700 hPa. The more opaque the blue cross, the more important is the specific difference according to the random forest feature importance. The higher the value $\Delta\theta_{ver}$, the more stably stratified is the atmosphere at this point.

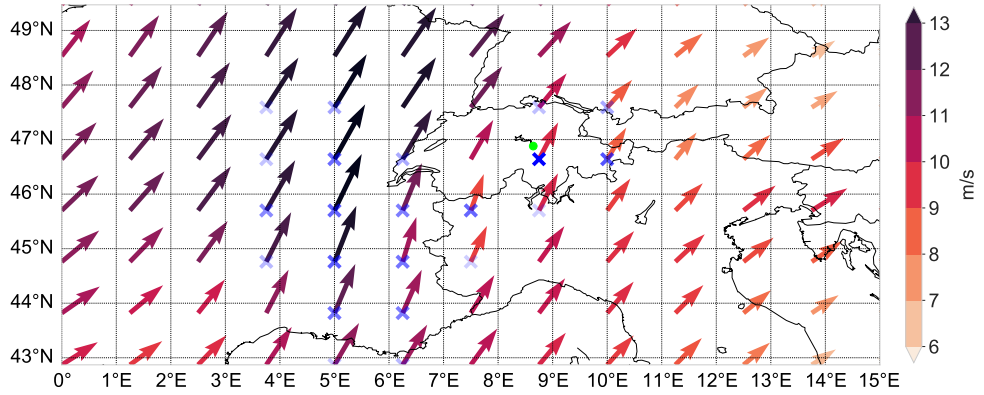


Figure 5.5: Same as Figure 5.4 but for U and V on 700 hPa.

collection of most important features, however, only under the implications of their spatial correlation.

Foehn predictability on ERAI

Optimizing the hyperparameters via three-fold cross-validation on $\boldsymbol{D}_{train}$ finally yielded an XGBoost model which consisted of $B = 200$ and $d_{max} = 4$. Table 5.1 shows the resulting confusion matrix of the XGBoost model after adjusting the threshold to balance false positives and false negatives. The different metrics to evaluate model performance (for calculation see Subsection 2.2.5), can be seen in Table 5.2. The selected features and their importance (see Equation 2.28) in the final model are shown in Table 5.3.

Table 5.1: Confusion matrix for D_{test} with a threshold of $c = 0.91$.

		Predicted label $\hat{y}$		
		Foehn	No-foehn	Total
Observed label y	Foehn	605	165	770
	No-foehn	165	13601	13766
	Total	770	13766	14536

Table 5.2: Scoring results on D_{test} with a threshold of $c = 0.91$.

Accuracy	Precision	Recall	F1-Score	AUC	Log loss
0.977	0.786	0.786	0.786	0.991	0.099

Table 5.3: Shown are the most important features and their importance of the final XGBoost model fitted to D_{train}.

feature	location/s	height	importance
ΔSLP	45.7, 8.75 - 47.59, 8.75	sea level	0.3783
ΔZ	45.7, 11.25 - 45.7, 7.5	850 hPa	0.0744
$\Delta\theta_{hor}$	43.82, 3.75 - 47.59, 8.75	850 hPa	0.0568
ΔSLP	45.7, 11.25 - 47.59, 8.75	sea level	0.0545
ΔSLP	45.7, 10.0 - 47.59, 8.75	sea level	0.0190
ΔZ	45.7, 12.5 - 46.64, 7.5	850 hPa	0.0168
$\Delta\theta_{hor}$	45.7, 11.25 - 47.59, 10.0	850 hPa	0.0144
ΔSLP	46.64, 11.25 - 47.59, 10.0	sea level	0.0132
ΔZ	46.64, 15.0 - 46.64, 8.75	700 hPa	0.0129
$\Delta\theta_{hor}$	46.64, 10.0 - 47.59, 11.25	850 hPa	0.0118

Generalization to CESM

After conducting the fitting process as described in Section 4.1, we obtained the following metrics on D_{test} (see Table 5.4). Here, the metrics were calculated after aggregating the predictions for all months.

Then, we went on to apply the model on CESM-p and CESM-f data. Since we wanted to ensure that the model actually captures the synoptic condition for foehn, we visualized composite plots in the same manner as for the ERAI (compare Figure 5.1 to Figure 5.5). Only this time, we used samples for which the algorithms predicted foehn. Exemplary, the results for the geopotential height in CESM-p are shown in Figure 5.6 and for CESM-f in Figure 5.7. For readability, the plots for all variables were moved to Appendix B.

Table 5.4: Scoring results on D_{test} under constraint optimization.

Accuracy	Precision	Recall	F1-Score	AUC	Log loss
0.975	0.766	0.766	0.766	0.988	0.067

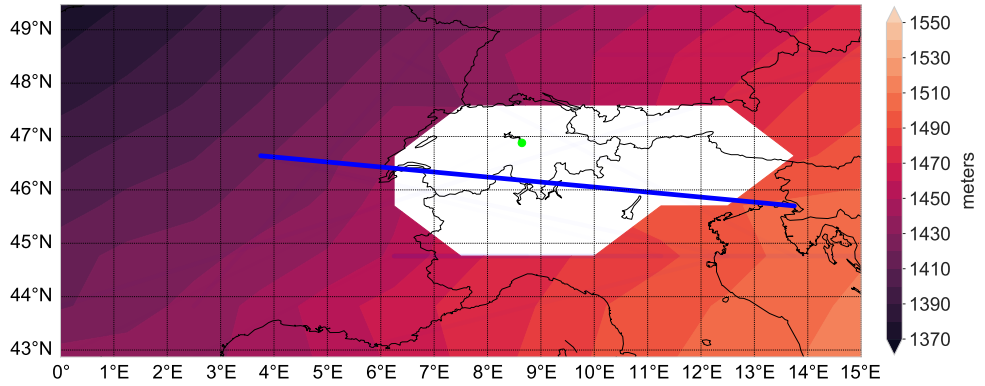


Figure 5.6: Shown are the most important geopotential height differences ΔZ over the mean Z weather condition on 850 hPa in CESM-p for all predicted foehn events. The more opaque the blue line, the more important is the specific difference. Due to the large feature importance of one feature, the other ones are barely visible. We kept this scaling to underline the fact that one feature appears to be enough to identify south foehn to the largest extent. The green dot marks the location of Altdorf.

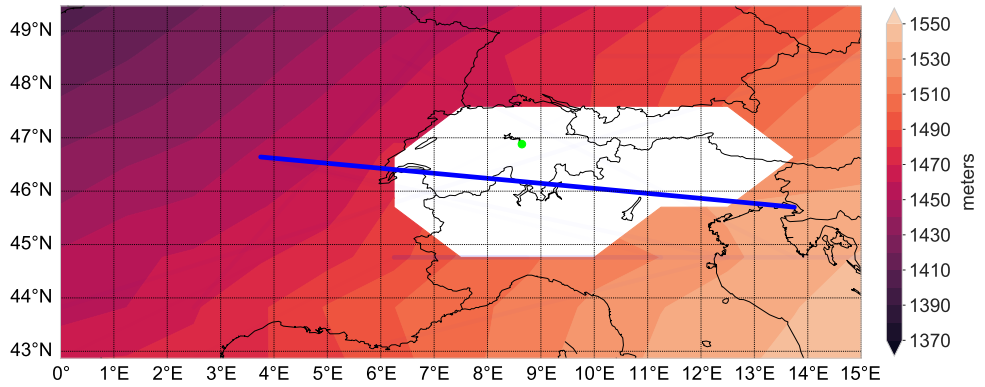


Figure 5.7: Same as Figure 5.6 but for CESM-f.

Comparison of monthly foehn frequencies

First, we compared the frequency of foehn in each dataset. Note that frequency was defined as the number of samples which showed/predicted foehn divided by the number of all samples within a specified data range (i.e., overall or in a specific month)

$$f = \frac{\sum_{y^{(i)}=1}}{\sum_{y^{(i)}}} \quad \text{or} \quad f = \frac{\sum_{\hat{y}^{(i)}=1}}{\sum_{\hat{y}^{(i)}}} \tag{5.1}$$

The overall frequencies of foehn in the observational, predicted ERAI, CESM-p, and CESM-f datasets were 5.4%, 5.4%, 5.1%, and 5.1%, respectively. Here, for CESM, this number resembled the mean frequency averaged over all ensemble members.

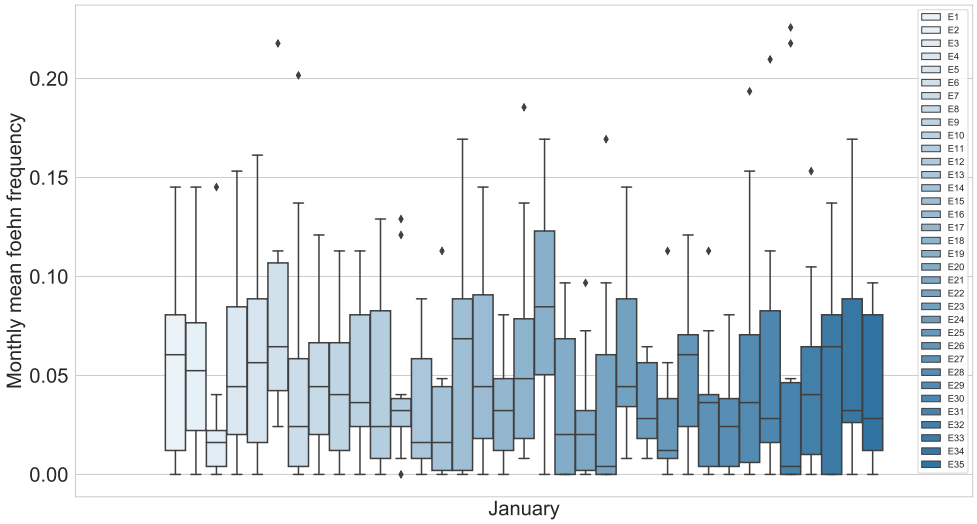


Figure 5.8: Shown is the distribution of predictions for all 35 CESM-p ensemble members for the month January.

Second, we visualized the foehn frequency for different ensembles members in CESM since we were interested if foehn differs between them. In Figure 5.8, this is exemplarily shown for CESM-p in the month January.

Finally, Figure 5.9 summarizes the monthly frequencies of all datasets. Here, one data point resembled one foehn frequency within a month grouped by year, and additionally, for CESM, by ensemble member. This resulted in ten data points for observational data and $\boldsymbol{D}_{test}$ (one for each year), and 350 data points for CESM-p and CESM-f (one for each year in each ensemble member). We did observe deviations, which will be evaluated for statistical significance in the next step.

Remember that we are going to follow the three steps which were described in Section 4.1. The results for the first test were calculated from the confusion matrix for every month and can be seen in Figure 5.10. Performing the second test yielded Figure 5.11. The third test (Wilcoxon rank-sum significance test) allowed us to calculate the p-values, which can be seen in Table 5.5.

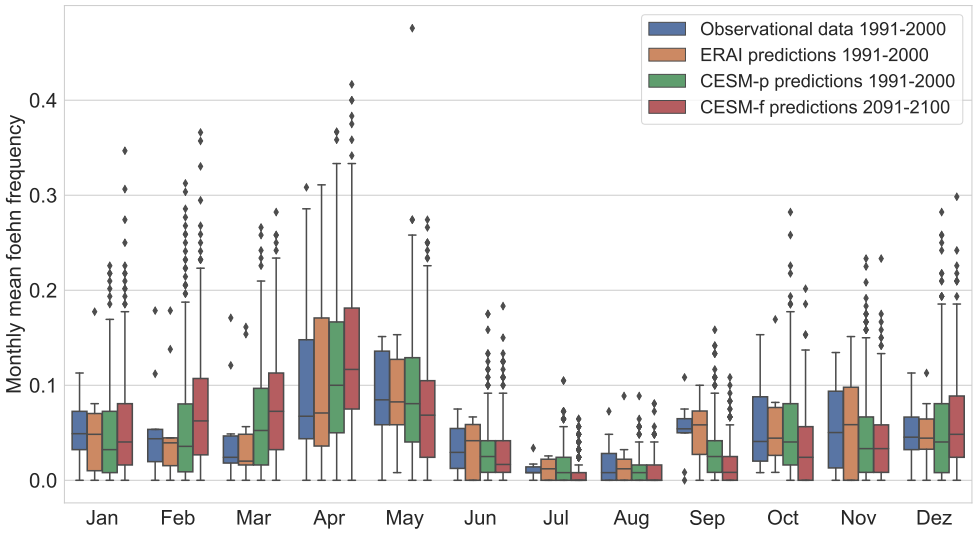


Figure 5.9: Shown are the mean observed/predicted monthly south foehn frequencies. First, the observational data of the period 1991-2000 is shown in blue. Second, the predictions of the XGBoost model for $\boldsymbol{D}_{test}$ are shown in yellow. Third, the predictions for all 35 ensemble members in CESM-p are shown in green. Lastly, the predictions for all 35 ensemble members for CESM-f are shown in red.

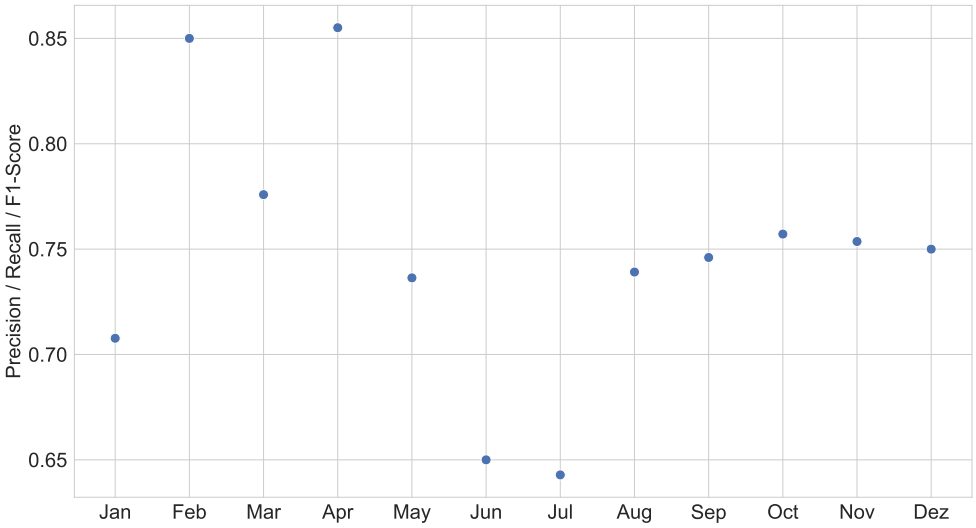


Figure 5.10: Shown are $Recall$, $Precision$ and $F1$-score on a monthly resolution. For each month the threshold was adjusted to balance false positives and false negatives.

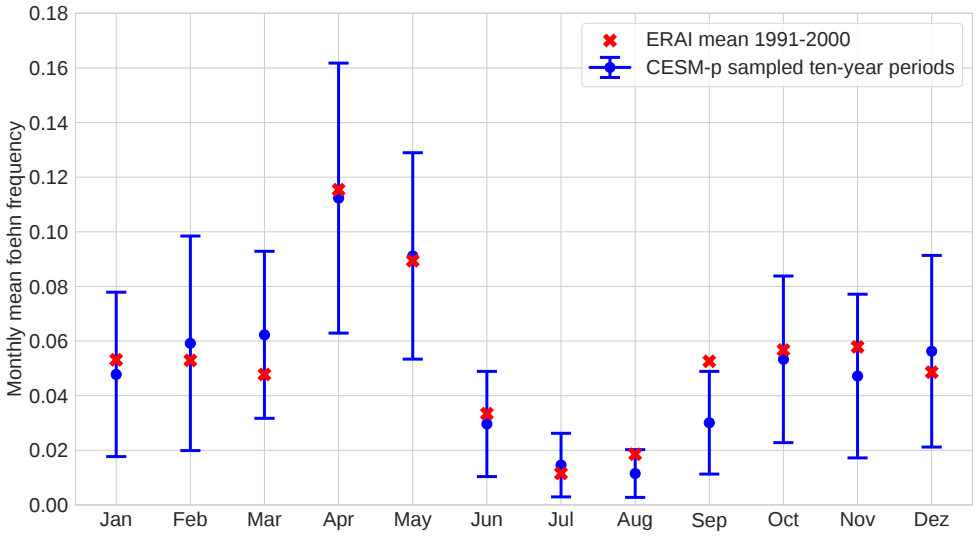


Figure 5.11: Shown are the predicted foehn frequency values aggregated over ten years for ERAI and CESM-p data. Here, we sampled ten-year periods of CESM-p as described in Section 4.1 and plotted the resulting Gaussian distribution together with mean and second standard deviation.

Table 5.5: Shown are the p-values from the Wilcoxon rank-sum test for different months and their significance under the global $\alpha = 0.05$. Note that we applied the Bonferroni correction due to multiple testing.

Month	p-value	Significant?
Jan	0.07539	No
Feb	0.0	Yes
Mar	0.00024	Yes
Apr	0.0026	Yes
May	0.00074	Yes
Jun	0.52486	No
Jul	0.0	Yes
Aug	0.01904	No
Sep	0.0	Yes
Oct	0.0	Yes
Nov	0.5774	No
Dez	0.08729	No

North foehn development

The selected most important features for north foehn are shown in Figure 5.12 until Figure 5.16.

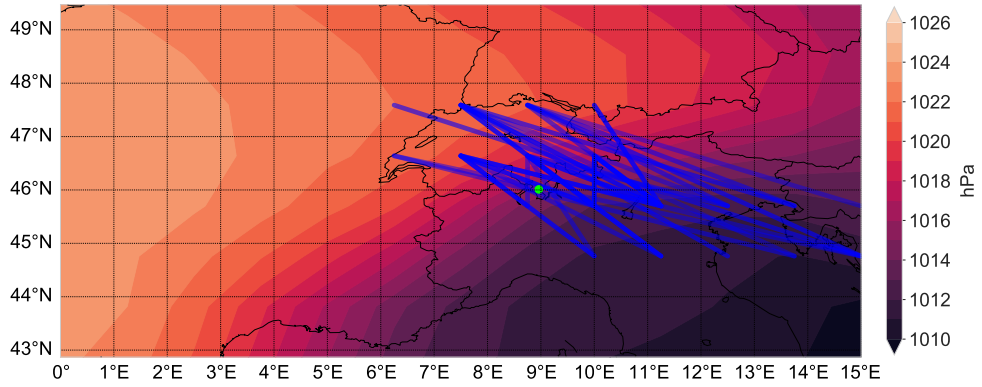


Figure 5.12: Same as Figure 5.1 but for ΔSLP for north foehn.

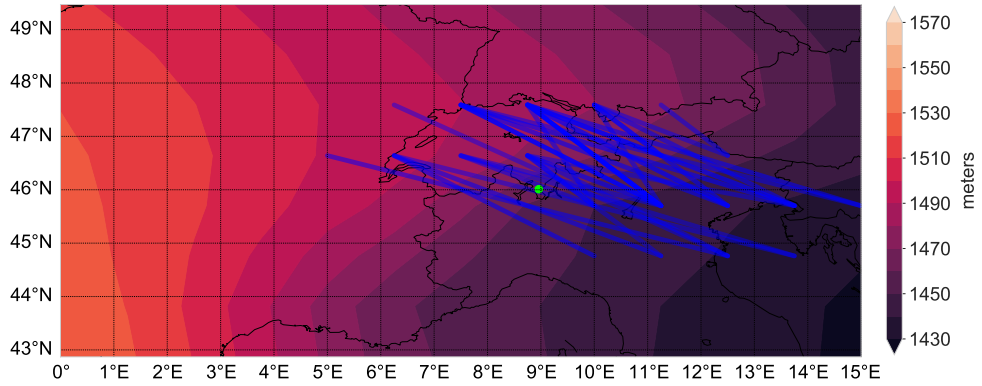


Figure 5.13: Same as Figure 5.2 but for ΔZ on 850 hPa for north foehn.

Optimizing the hyperparameters via three-fold cross-validation on $\boldsymbol{D}_{train}$ finally yielded a XGBoost model with $B = 250$ and $d_{max} = 4$. Table 5.6 shows the resulting confusion matrix of the XGBoost model after adjusting the threshold to balance false positives and false negatives. The different metrics to evaluate model performance can be seen in Table 5.7. The ten most important features are shown in Table 5.8.

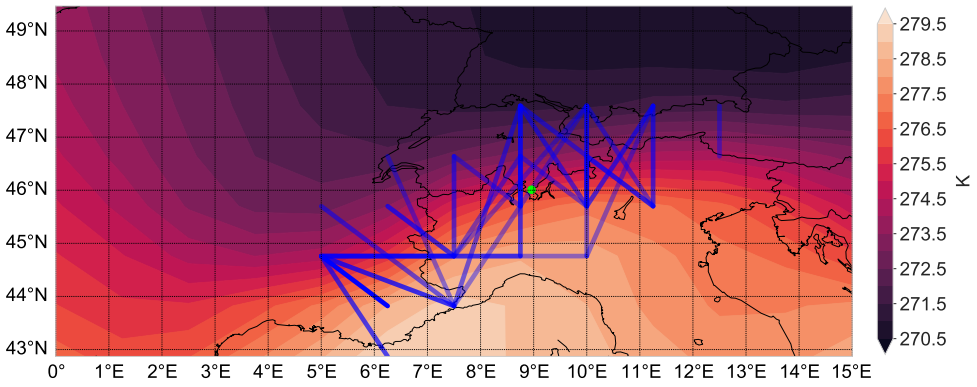


Figure 5.14: Same as Figure 5.3 but for $\Delta\theta_{hor}$ on 850 hPa for north foehn.

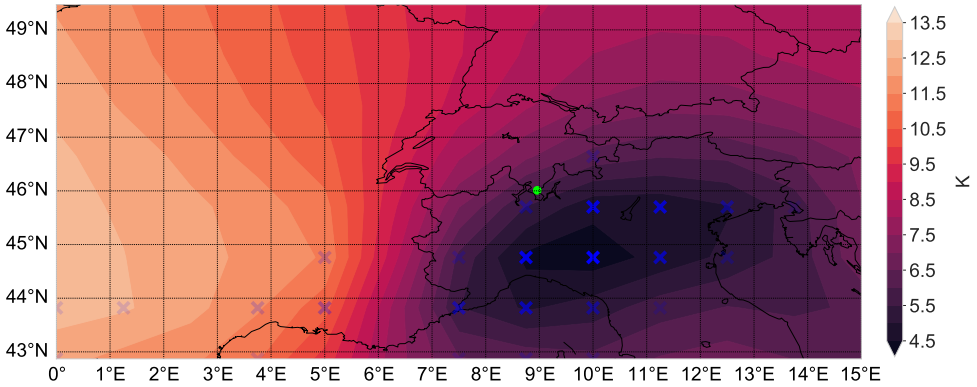


Figure 5.15: Same as Figure 5.4 but for $\Delta\theta_{ver}$ between 900 and 700 hPa for north foehn.

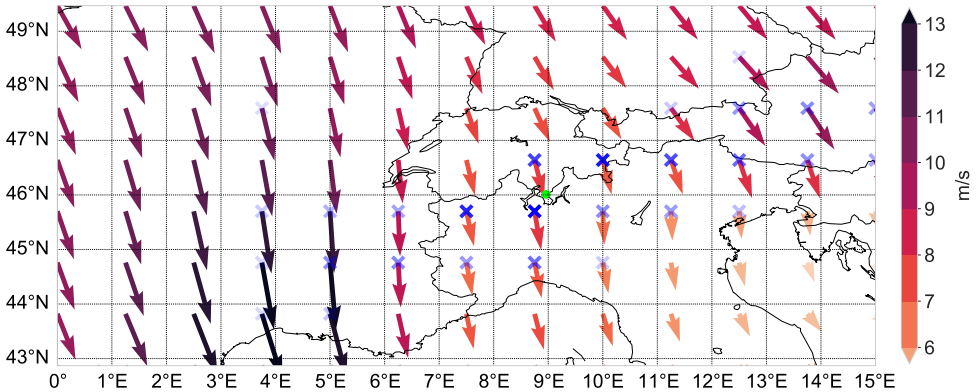


Figure 5.16: Same as Figure 5.5 but for U and V on 700 hPa for north foehn.

Table 5.6: Same as Table 5.1 but for north foehn with $c = 0.92$.

		Predicted label $\hat{y}$		
		Foehn	No-foehn	Total
Observed label y	Foehn	734	208	942
	No-foehn	208	13412	13620
	Total	942	13620	14562

Table 5.7: Same as Table 5.2 but for north foehn with $c = 0.92$.

Accuracy	Precision	Recall	F1-Score	AUC	Log loss
0.972	0.779	0.779	0.779	0.989	0.129

Table 5.8: Same as Table 5.3 but for north foehn.

feature	location/s	height	importance
ΔZ	45.7, 10.0 - 46.64, 8.75	850 hPa	0.4166
ΔSLP	45.7, 11.25 - 46.64, 8.75	sea level	0.0926
ΔSLP	45.7, 10.0 - 46.64, 8.75	sea level	0.0457
ΔZ	45.7, 11.25 - 46.64, 8.75	850 hPa	0.0340
ΔZ	47.59, 13.75 - 47.59, 5.0	700 hPa	0.0306
ΔSLP	45.7, 10.0 - 47.59, 8.75	sea level	0.0249
ΔZ	45.7, 13.75 - 46.64, 8.75	850 hPa	0.0091
ΔZ	43.82, 6.25 - 44.76, 6.25	700 hPa	0.0081
$\Delta\theta_{hor}$	43.82, 0.0 - 48.53, 5.0	850 hPa	0.0078
ΔZ	44.76, 13.75 - 46.64, 6.25	850 hPa	0.0075

The metrics on D_{test} after the constraint optimization are shown in Table 5.9. The Z composite plots for predictions on CESM-p and CESM-f can be seen in Figure 5.17 and Figure 5.18. Again, the remaining composite plots have been moved to Appendix B.

Table 5.9: Same as Table 5.4 but for north foehn.

Accuracy	Precision	Recall	F1-Score	AUC	Log loss
0.963	0.714	0.714	0.714	0.979	0.094

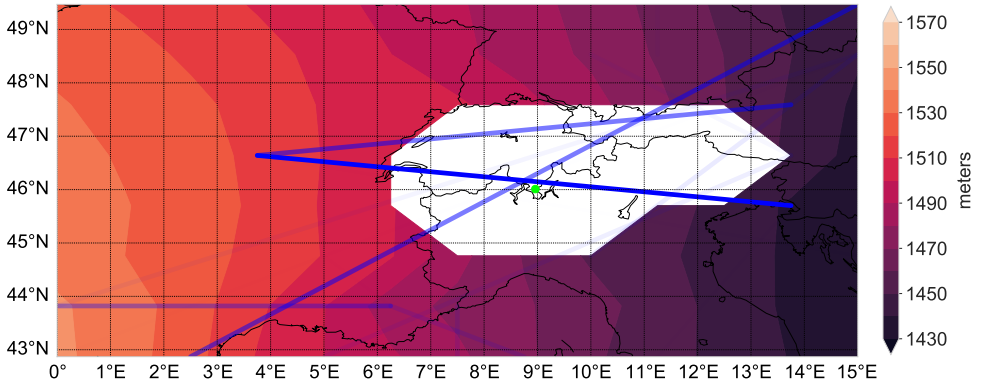


Figure 5.17: Same as Figure 5.6 for ΔZ on 850 hPa in CESM-p but for north foehn.

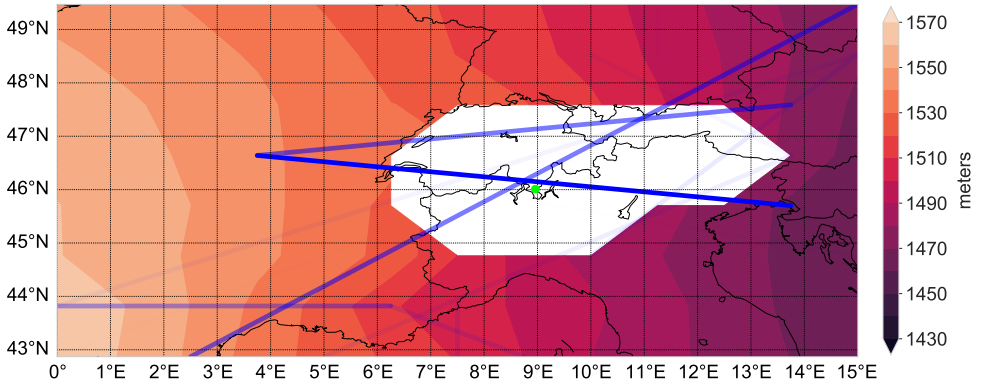


Figure 5.18: Same as Figure 5.7 for ΔZ on 850 hPa in CESM-f but for north foehn.

The frequencies of foehn in the observational, predicted ERAI, CESM-p, and CESM-f data were 6.4%, 6.5%, 6.0%, and 5.9%, respectively. The ensemble members still showed a large variability for north foehn (see Figure 5.19). The final comparison for ERAI, CESM-p, and CESM-f can be obtained from Figure 5.20.

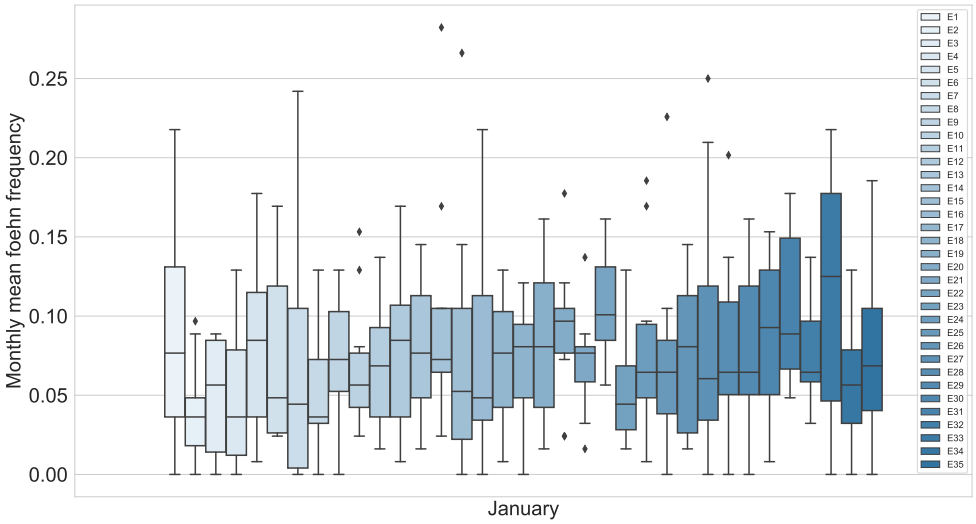


Figure 5.19: Same as Figure 5.8 but for north foehn.

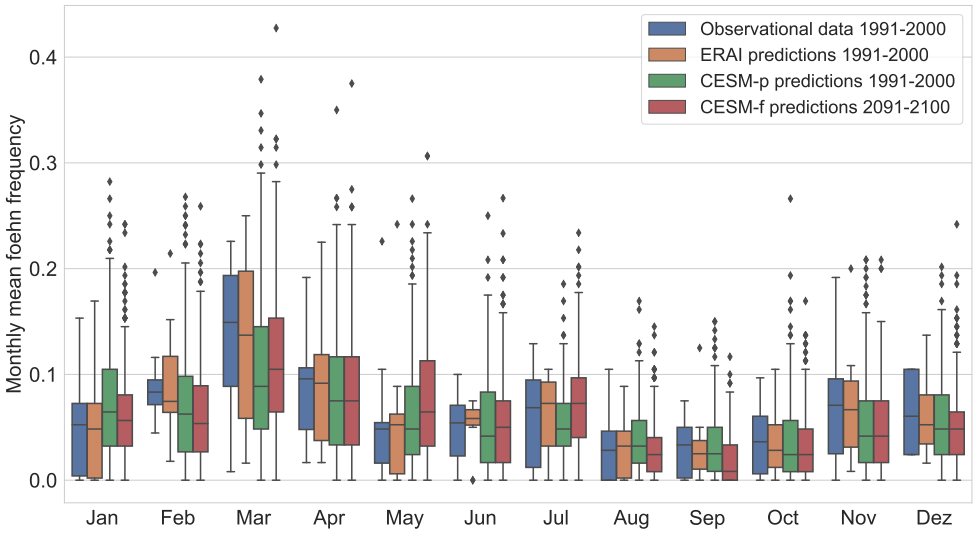


Figure 5.20: Same as Figure 5.9 but for north foehn.

The monthly scores on D_{test} for north foehn can be seen in Figure 5.21. The results of the statistical test, whether ERAI could be seen as a potential CESM-p member, are shown in Figure 5.22. The results of the Wilcoxon rank-sum test and the corresponding p-values can be found in Table 5.10.

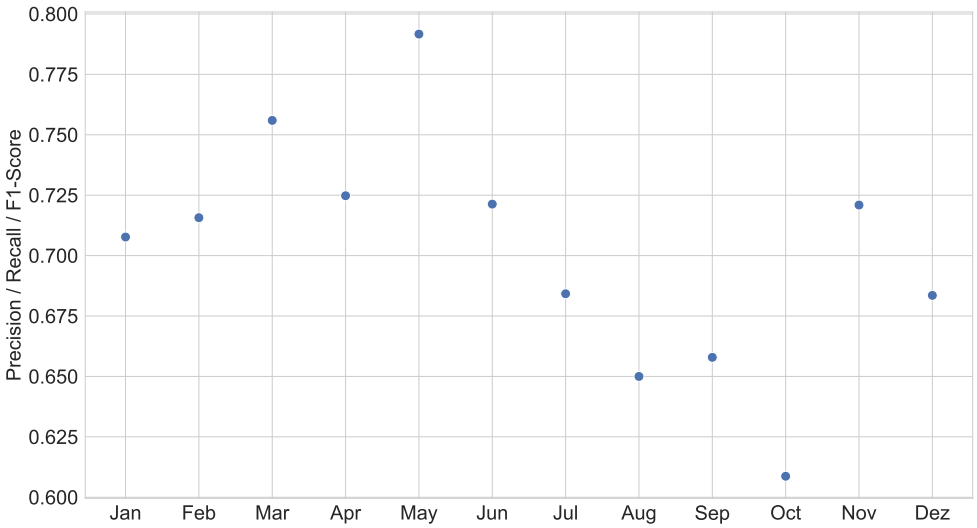


Figure 5.21: Same as Figure 5.10 but for north foehn.

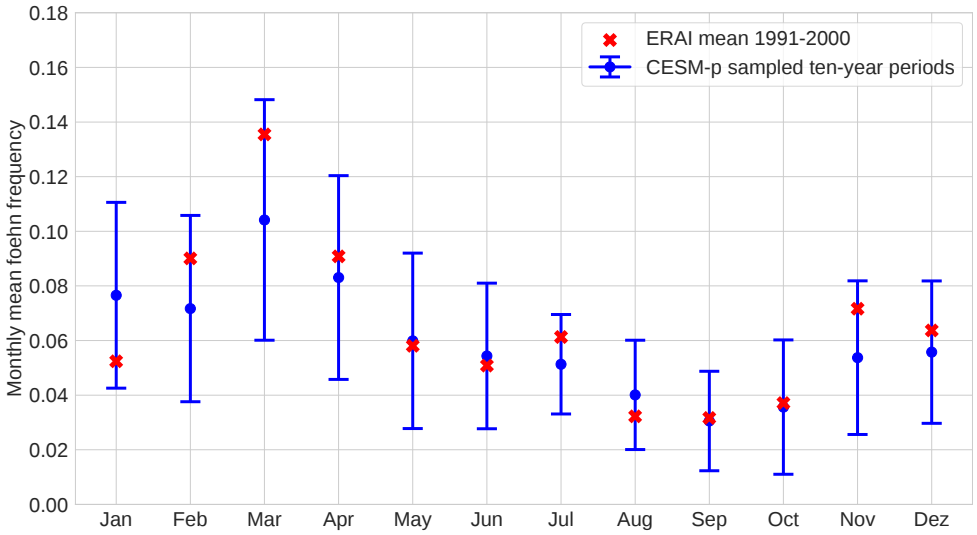


Figure 5.22: Sames as Figure 5.11 but for north foehn.

Table 5.10: Same as Table 5.5 but for north foehn.

Month	p-value	Significant?
Jan	0.00069	Yes
Feb	0.1142	No
Mar	0.01063	No
Apr	0.93475	No
May	0.00023	Yes
Jun	0.80469	No
Jul	0.0	Yes
Aug	0.00013	Yes
Sep	0.0	Yes
Oct	0.30088	No
Nov	0.28967	No
Dez	0.05184	No

5.1.2 Discussion

Synoptic-scale foehn conditions

The identification of the most important features yielded satisfactory and physically meaningful results for the different variable categories. At first, the random forest selected features that laid in the Alpine region (and not somewhere far from it) and thus coincided with a naive expectation. In the next paragraphs, we look into the different selected features in detail.

For the sea level pressure (see Figure 5.1), we observed a high-pressure system over northern Italy and a low-pressure system to the north-west of the Alps. Moreover, the typical "foehn knee" in the isobars is discernible (Richner and Hächler, 2013; Sprenger et al., 2016). The selected ΔSLP differences laid in the region of the largest gradient. This finding agrees with the literature (Drechsel and Mayr, 2008; Richner and Hächler, 2013; Plavcan et al., 2014) that a strong pressure gradient is a necessary but not sufficient condition for foehn. Next, we compared the selected pressure features to the pressure component of the established Widmer index. However note that the Widmer index has a slightly different purpose - namely forecasting. The pressure component of the Widmer index looks at the pressure difference between Venice and Tours (Courvoisier and Gutermann, 1971). In contrast to this difference, we found that foehn can be diagnosed from even more local features.

Second, the composite plot for the geopotential height Z on 850 hPa (see Figure 5.2) showed a similar picture as the sea level pressure. As Gerstgrasser (2017) noted, geopotential height differences are a suitable predictor for foehn strength (especially for wind speed maxima). Again, the model selected ΔZ's, which showed a pronounced difference for foehn events. Interestingly, if we compared the selected features against the geopotential height component of the Widmer index, which is the difference between Rijeka and Valence (Courvoisier and Gutermann, 1971), we found that the selected features align well with this component.

Third, the potential temperature θ_{hor} on 850 hPa showed an interesting pattern in such a way that the air is on average up to 5 K potentially warmer to the north of the Alps compared to the south of the Alps. This pattern is also observed and described in Gerstgrasser (2017) and Hächler et al. (2011). Advection of warm air north to the Alps and the existence of cold air to the south of the Alps is typical for deep foehn events. The difference in temperature between the two air masses causes a local hydrostatic pressure gradient. Apparently, these are also the differences the model considered.

Forth, we investigated the stability of the atmosphere through the vertical potential temperature difference $\Delta\theta_{ver}$ between 700 hPa and 900 hPa. The larger this difference, the more stably stratified is the atmosphere. In Fig-

ure 5.4, we observed a stably stratified atmosphere to the south of the Alps and a more unstable atmosphere to the north of the Alps. As discussed in Steinacker (2006), the stable stratification is typical for Austrian foehn, when air is blocked orographically to the south of the Alps. Apparently, a more unstably stratified atmosphere to the north of the Alps is also typical during foehn.

Lastly, the mean wind field on 700 hPa during foehn is depicted in Figure 5.5. One can observe the typical southerly/southwesterly flow over the Alps (Gerstgrasser, 2017). In our case, the most important wind features were identified from the meridional flow V.

In conclusion, we can say that the most important features which characterize the large-scale fingerprint of foehn can be identified from the synoptic conditions over the Alps in reanalysis data and observational foehn data. Here, the identified features had a strong physical interpretation and reflected the current knowledge about foehn. In the next section, we assessed which skill foehn can be identified from ERAI data.

Foehn predictability on ERAI

By looking at the metrics, the final XGBoost model can be compared to previous work of other scientists who predicted or nowcasted foehn with machine learning models (Sprenger et al., 2017; Zweifel et al., 2016). The comparison can be obtained from Table 5.11. Note that the raw values are only comparable to some extent due to the different purposes of the other projects. Nevertheless, being aware of this caveat, we see that our model showed similar if not increased performance compared to prior models. This is remarkable because our model is based on very coarse ERAI data with a spatial resolution of approximately 80 km. In contrast, the COSMO-7, COSMO-2, and ECMWF models have a resolution of 7 km, 2.2 km, and 16 km, respectively. Nevertheless, we managed to achieve similar scores. That also supports the claim from Drechsel and Mayr (2008) and Plavcan et al. (2014) that foehn is, in general, predictable from its synoptic fingerprint in NWP data, as long as the obstacle is resembled sufficiently.

Let us now consider the most important features selected by the model in Table 5.3. As one would expect from Sprenger et al. (2017), Zweifel et al. (2016), and Gerstgrasser (2017), the XGBoost model did select pressure and geopotential height differences as its most important features. In addition, also the hydrostatic gradient induced by the difference in potential temperature discussed in Gerstgrasser (2017) appears to have played an important role. Including these synoptic-scale features might enhance the performance of other models in the future.

Hence, we can conclude the second objective of this thesis: In general, it is

Table 5.11: Score comparison between different models from previous scientific work. All algorithms were applied to the location of Altdorf. Sprenger et al. (2017) used an AdaBoost algorithm to nowcast foehn based upon the COSMO-7 model. Zweifel et al. (2016) used logistic regression with L1-regularization to predict foehn on ECMWF and COSMO-2 data. However note that they forecasted foehn 15 hours in advance. Furthermore, they also used observational data from several weather stations in addition to the NWP model data.

	Accuracy	Precision	Recall	F1-Score	AUC
Our model	0.977	0.786	0.786	0.786	0.991
AdaBoost	0.965	0.662	0.882	0.756	NA
Log. Regr. COSMO-2	0.960	0.800	0.627	0.703	0.974
Log. Regr. ECMWF	0.969	0.841	0.725	0.779	0.986

possible to infer the existence of foehn from synoptic conditions on a grid as coarse as 80 km. The skill is highly comparable to using NWP models with higher resolution.

Generalization to CESM

Now, we discuss the transition from ERAI to CESM data. From Table 5.4 we obtain that we only lost little skill compared to the baseline model that was fitted solely on ERAI (see Table 5.2). Interestingly, the log loss even decreased. This could be due to the fact that the model now predicted less samples completely off (e.g., $y^{(i)} = 1$ but $p^{(i)} \approx 0$) since these are the mistakes the log loss punishes the harshest. For this reason, we further compared the composite maps for foehn cases in ERAI (Figure 5.2), CESM-p (Figure 5.6) and CESM-f (Figure 5.7), here only exemplary for Z. In all three cases, we were able to observe the typical foehn pattern that we have discussed above. The same holds for the other variables whose plots can be found in Appendix B.

One fact that could surprise is that we reconstructed the whole weather conditions for foehn over Europe well, even though the model appeared to mainly look at a single feature within a variable category. This suggests that the synoptic foehn situation is in principle identifiable from a set of few, well-selected parameters only. Further, this agrees with the findings of Widmer (1966) and Courvoisier and Gutermann (1971), who also achieved high foehn predictability from four and two parameters, respectively.

We conclude that, in principle, it is possible to transition from ERAI data to CESM data and identify the same synoptic foehn situation in both models. Surprisingly, for south foehn, only very few, well-selected parameters sufficed to capture the whole synoptic situation.

Comparison of monthly foehn frequencies

First, we compared the raw frequency of foehn between ERAI and CESM-p. With foehn being observed on average 5.1% of the time in CESM-p, for the same period, we found 5.4% in ERAI. As we argue later also for the monthly data, ERAI data resembles a potential realization of the weather in the period 1991 to 2000. CESM-p spans a space of potential weather realizations (i.e., 35 of them). Thus, we compared the means of each ensemble member against the ERAI mean frequency. We found that the ERAI data laid at the 75% quantile of the foehn frequency distribution from the CESM-p ensemble members. Therefore, we argue that the ERAI frequency is within the range of what has also been simulated by CESM-p. However, we do not want to exclude other sources of error like small biases in producing enough foehn situations due to the different topography between ERAI and CESM (compare Figure 3.2 and Figure 3.3).

Second, also notable is the spread between the ensemble members for specific months, as shown in Figure 5.8. Depending on which ensemble member one looks at, a substantially varying foehn frequency could be observed. Due to this variability, we will only be able to make statistical conclusions for all ensemble members combined.

Third, investigating the monthly distribution of foehn in Figure 5.9, we observed that we approximated the annual cycle of foehn in CESM-p quite well despite the large variance between ensemble members. However, some larger deviations between observational data and CESM-p could be spotted in September. Also notable is that in spring (FMA), the CESM-p and CESM-f scenarios showed some larger differences, i.e., during these months more years in more ensemble members showed a larger foehn frequency in CESM-f. In the next step, we controlled whether those differences were significant. Here, we followed the structure outlined in Section 4.1.

1. **Observational data vs. ERAI predictions.** First, from Figure 5.10 we obtained that prediction worked best during spring, autumn, and winter. Here, our skill ranged from 0.71 to 0.85 as measured by the *Precision*, *Recall*, and $F1$-score. During summer (JJ), our prediction skill was worst, residing around 0.65 in *Precision*, *Recall*, and $F1$-score. These findings also agree with Sprenger et al. (2017), who found the best predictability in winter/spring and the worst during summer. Further, they explained that during summer, mechanisms like increased solar irradiance lead to a local heat low in the Alpine region. The resulting pressure field and thermally-driven valley circulation counteracts the foehn flow (Lotteraner, 2009). Consequently, this may have lead to a loss in predictive power of the features the model has learned and explains the reduced scores.

Note that each month was treated with a different threshold such that all three metrics coincided. The idea of applying a different threshold for every month to compensate for potential external factors the model fails to capture has also already been applied in the Widmer index (Widmer, 1966).

We conclude with the result that foehn is predictable from coarse NWP data with varying skill over the months. In most months, we found good predictability, while it was worst during summer. However, in principle, each false negative should be compensated by a false positive and hence not affect the final frequency. Also, it should be noted that we still were far from random prediction. In this case, we would find $Precision = Recall \approx 0.02$, which is the true observed frequency during summer. Nevertheless, since in June and July we had very few positive samples (approximately 15 on $\boldsymbol{D}_{test}$), we rejected to trust in the model during these months.

2. **ERAI predictions vs. CESM-p predictions.** In Figure 5.11, we compared how likely it would be to observe the same predictions during a month in a CESM-p ensemble member as in ERAI. Here, we relied on the second standard deviation of the sampled Gaussian distribution. For this reason, the ERAI representation laying outside these intervals would only happen by a random chance of 5%. Under these circumstances, a transition from ERAI to CESM-p would be rejected since apparently, the model failed to generalize well. In our case, this only happened for September. In all other cases, ERAI coincided with what one would expect under the CESM-p ensemble. Consequently, we only rejected a viable transition from ERAI to CESM-p for September. Note however that the random chance for one or more months falling outside the second standard deviation for twelve months is at 46% as calculated with the Bernoulli formula.

3. **CESM-p predictions vs. CESM-f predictions.** Lastly, we used a standard Wilcoxon rank-sum test with Bonferroni correction to check whether the samples from CESM-p and CESM-f followed the same distribution. As we conclude from Table 5.5, we found a statistical significant difference for the months February, March, April, May, July, September, and October.

Putting everything together left us with the following conclusion for the fourth objective: We observed a significant increase in the monthly mean foehn frequency from the CESM-p to the CESM-f ensemble for the winter/spring months February, March, and April. Furthermore, for May and October, we found a significant decrease in foehn frequency. Which ensemble member will become a reality we can not say since we only can make a statistical estimate for all ensemble members combined. Thus, considering

Table 5.12: Score comparison between our model and the model from Zweifel et al. (2016), who used logistic regression with L1-regularization to predict north foehn on ECMWF and COSMO-2 data in Piotta. However note that they forecasted foehn 15 hours in advance and also used observational data from several weather stations in addition to the NWP model data. One can clearly see the misleading effect of the more severe class imbalance in Lugano in the *Accuracy*.

	Accuracy	Precision	Recall	F1-Score	AUC
Our model	0.972	0.779	0.779	0.779	0.989
Log. Regr. COSMO-2	0.904	0.854	0.824	0.839	0.961
Log. Regr. ECMWF	0.912	0.864	0.839	0.851	0.960

all ensemble members, we expect foehn to become more common during the late-winter/early-spring months and less common during May and October.

North foehn development

First, we briefly looked at the most important features in ERAI from Figure 5.12 to Figure 5.16. As we can obtain from the composite plots, the model again considered physical relevant features to identify north foehn from the synoptic situation. Again, the identified features coincided well with the synoptic situation described in previous literature (Cetti et al., 2015; Kljun et al., 2001).

When it comes to the model training and evaluation in Table 5.6 and Table 5.7, the results were highly comparable to those of south foehn. Again, the most important features in Table 5.8 were mainly geopotential height and pressure differences over the Alps. We tried to compare our model against other north foehn prediction models. Zweifel et al. (2016) attempted to predict north foehn in Piotta, which shows foehn occurrence more frequently than Lugano. Thus, once more, the results in Table 5.12 have to be taken with a grain of salt. Nevertheless, in the feature importances of their model, they also identified sea level pressure differences to be of great importance, which agreed with our findings.

When moving on to the CESM generalization, we observed that the skill dropped somewhat more substantially (compare Table 5.9 with Table 5.7). Thus, again we compared the different composite plots for ERAI, CESM-p and CESM-f (exemplary shown for Z in Figure 5.12, Figure 5.17, and Figure 5.18; see Appendix B for all plots). We see that the model mostly captured the typical conditions for north foehn. The slight differences are potentially linked to the different topography of the Alps in Figure 3.2 and Figure 3.3. Also, we recognized that apparently more features had similar importances to the models, which agrees with Zweifel et al. (2016), who also found that more parameters are needed to describe north foehn accurately. This could resemble an explanation for the decreased prediction scores.

Lastly, we captured the transition from ERAI to CESM for north foehn in

statistical terms. In total, we observe foehn 6.0% of the time in CESM-p compared to 6.5% in ERAI. Here, ERAI laid at the 90% percentile of the distribution from all CESM-p ensemble members, i.e., four members showed a larger foehn frequency. Thus, we found that ERAI could resemble one potential CESM-p member. When looking at the monthly frequencies in Figure 5.20 we again observed the intra-annual cycle from Figure 1.6. Next, we evaluated these distributions within each month by employing the statistical tests described in Section 4.1.

1. **Observational data vs. ERAI predictions.** First, from Figure 5.21 we obtained that in general on $\boldsymbol{D}_{test}$ we managed to capture the foehn quite well. For most months, the metrics were greater than 0.67. Only during the months August, September, and October they dropped to approximately between 0.61 and 0.66. Again, it could be that during those months, certain features lose their predictive power. However, each false negative should be compensated by a false positive and hence not affect the frequency. Furthermore, the models still showed some skill (> 0.6) with a reasonable amount of positive foehn samples (> 40). Thus, we did not reject any model for its suitability to diagnose foehn.
2. **ERAI predictions vs. CESM-p predictions.** In Figure 5.22, we see that we could accept the hypothesis that ERAI represents a possible realization of CESM-p for all months. Therefore, we say that our models predicted foehn on CESM-p reasonably well for each month.
3. **CESM-p predictions vs. CESM-f predictions.** Lastly, we used the Wilcoxon rank-sum test with Bonferroni correction also for the predicted distributions of north foehn on CESM-p and CESM-f. As we conclude from Table 5.10, we found significant differences for January, May, July, August, and September.

The conclusion left us with the following regarding the fifth objective: We observed a significant increase in the monthly mean foehn frequency from CESM-p to CESM-f for the months May and July. Additionally, for January, August, and September, we found a significant decrease in foehn frequency. Again, there is a large variance associated with the predictions between the ensemble members, thus we only make statistical statements for all ensemble members combined. For this reason, we expect north foehn to become more common during the two late-spring/summer months May and July and less common during January and the two late-summer/early-autumn months August and September.

5.2 Impact of foehn on forest fires

5.2.1 Results

Prior fire ignition

First, we investigated the 24-hour period. The resulting plot for the normalized count of fires can be seen in Figure 5.23. Furthermore, we contrasted non-foehn with foehn situations in Figure 5.24. For readability, we rescaled the y-axis with the minimum value for the following plots.

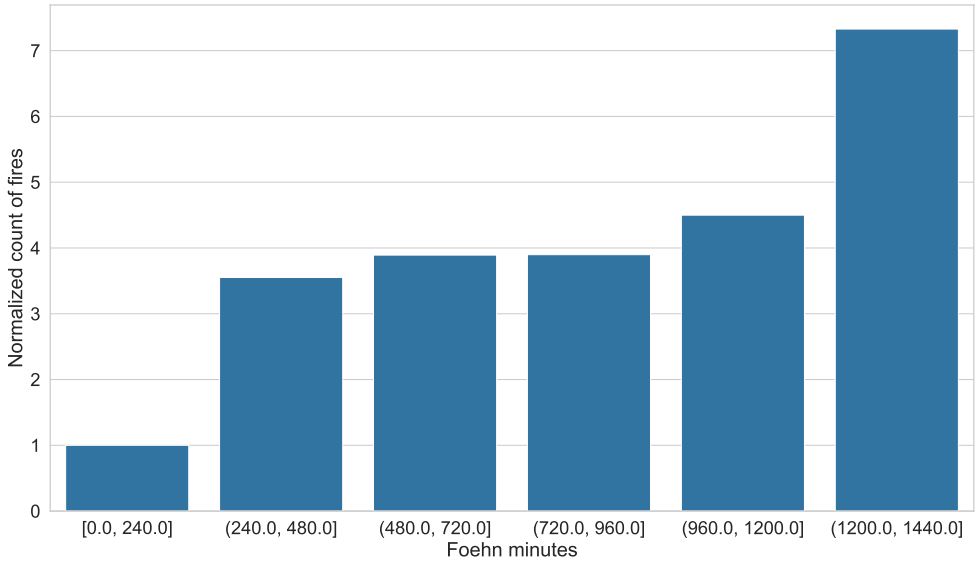


Figure 5.23: Normalized count of fires over the foehn minutes for the 24-hour period. As described in Section 4.2, we normalized the count of fires by the general occurrence of a certain foehn length.

The results regarding the effect of foehn temperature increase can be seen in Figure 5.25. Note that here, we only included fires associated with north foehn. For south foehn, the number of fires was too little to provide meaningful insight.

The results for the 48-hour period look similar and can be found in Appendix C.

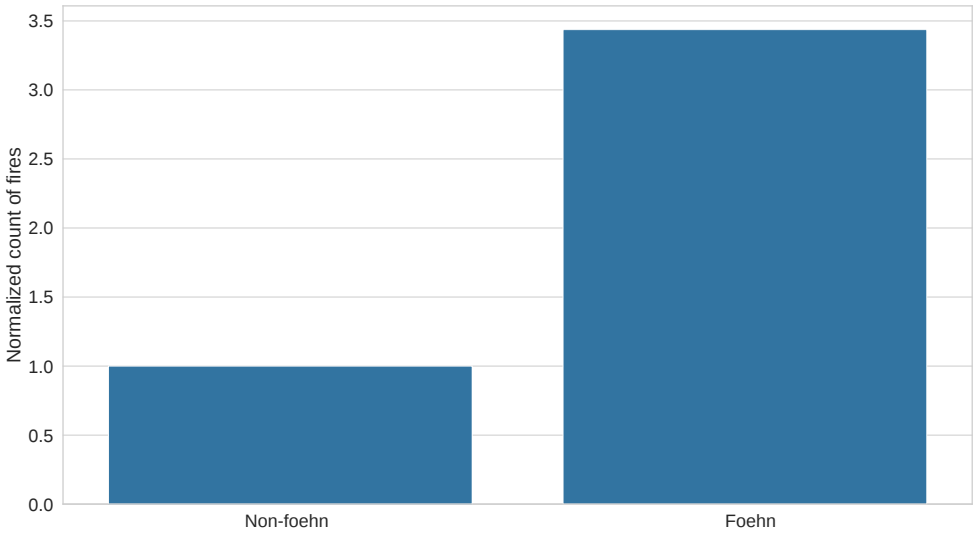


Figure 5.24: Normalized count of fires not linked (left) and linked (right) to foehn for the 24-hour period.

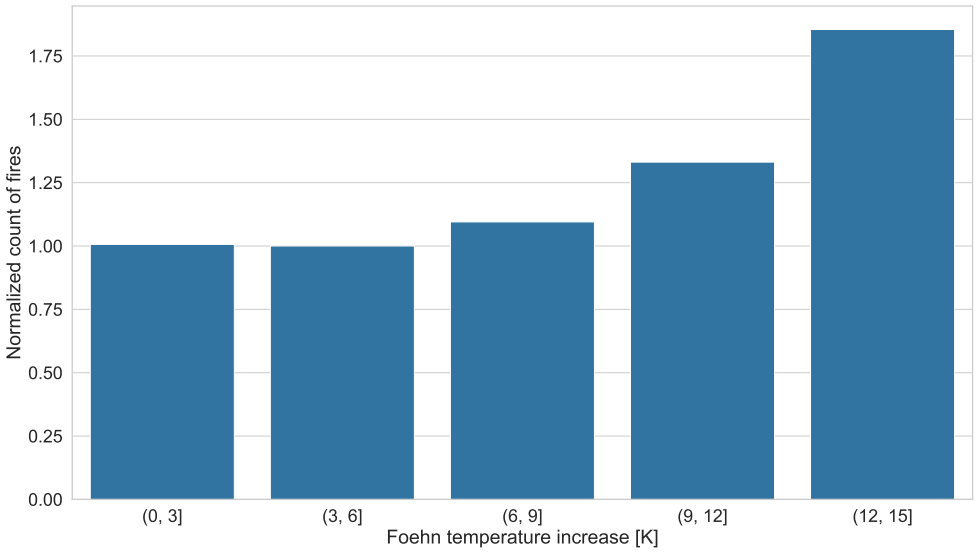


Figure 5.25: Normalized count of fires over the foehn temperature increase for the 24-hour period for north foehn fires. As described in Section 4.2, we normalized the count of fires by the general occurrence of a certain foehn temperature increase.

Post fire ignition

In the next step, we looked at the period six hours after fire ignition. In Figure 5.26, it is shown how foehn influenced the burned area. Again, we opposed non-foehn versus foehn fires in Figure 5.27. Please note the logarithmic scale.

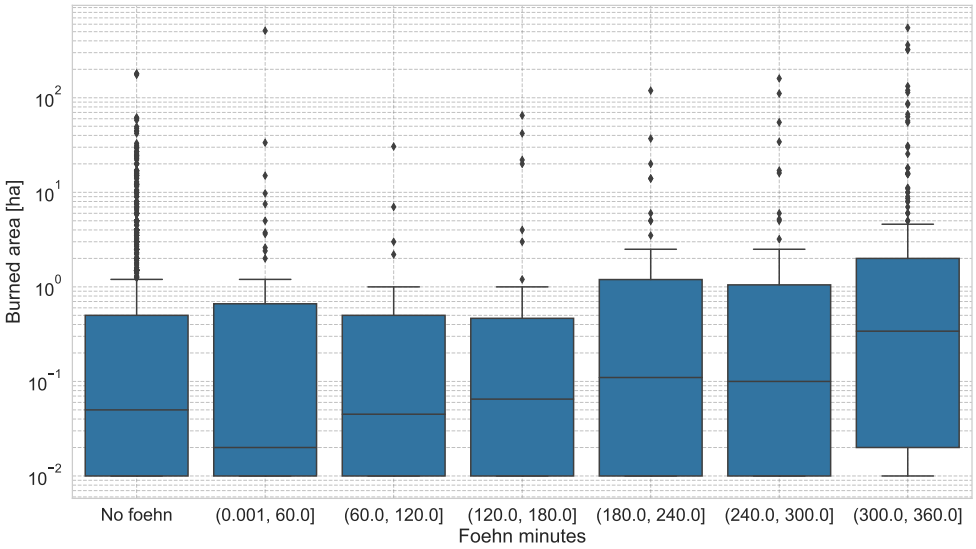


Figure 5.26: Burned area distribution over foehn minutes in the six hours after ignition.

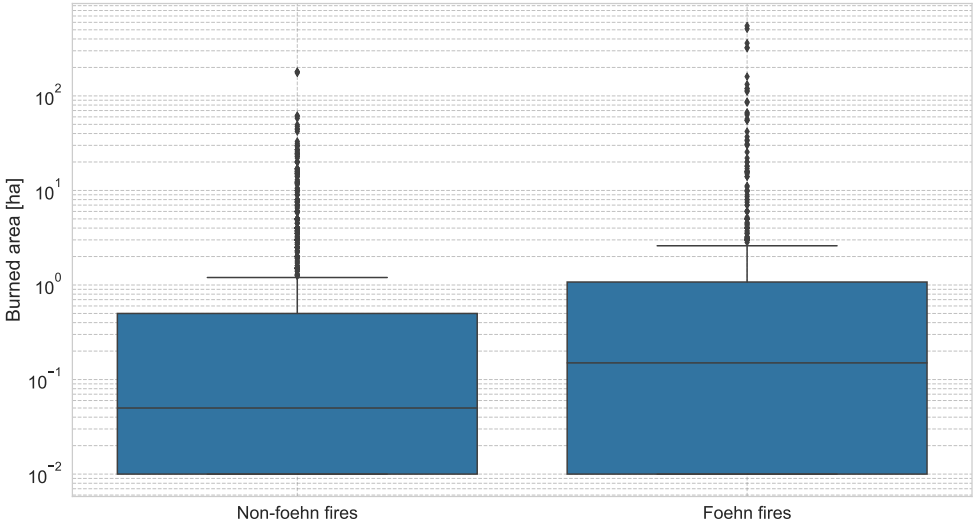


Figure 5.27: Burned area distribution for fires not linked (left) and linked (right) to foehn in the six hours after ignition.

Next, we investigated the data disaggregated by fire regime, foehn location, and decade. Since the results look similar to Figure 5.26, we only show the non-foehn versus foehn contrast plots here and moved the remainder to Appendix C.

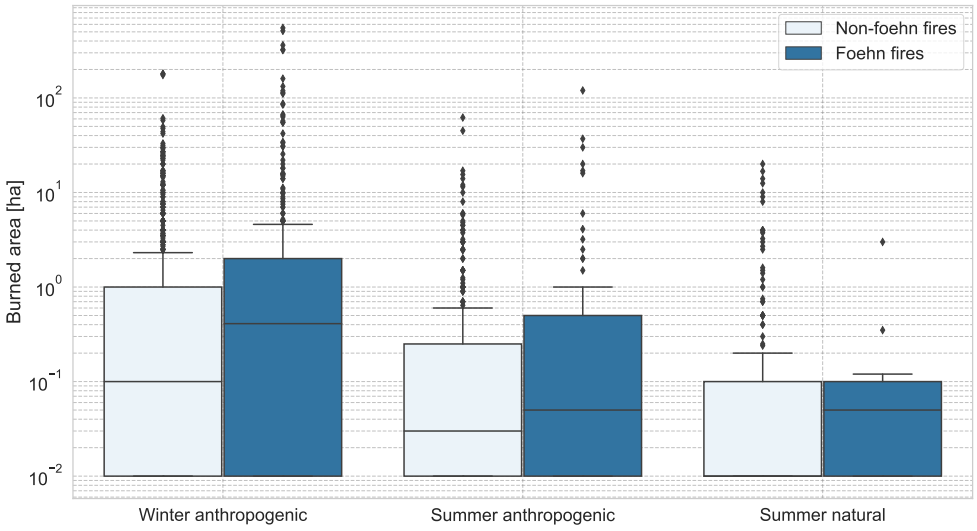


Figure 5.28: Burned area distribution for fires not linked (light-blue) and linked (dark-blue) to foehn in the six hours after ignition, here disaggregated by fire regime.

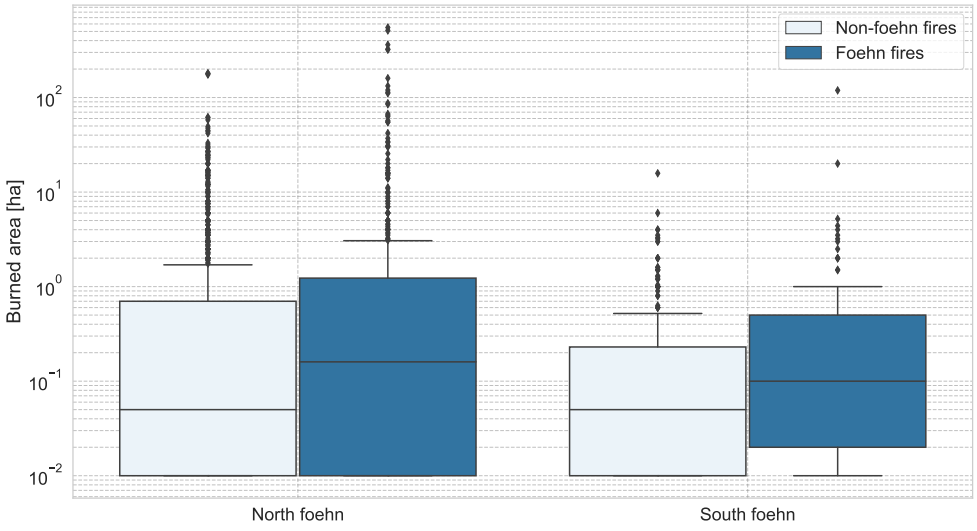


Figure 5.29: Same as Figure 5.28 but disaggregated by foehn location.

Table 5.13 shows the median burned area increases and their statistical sig-

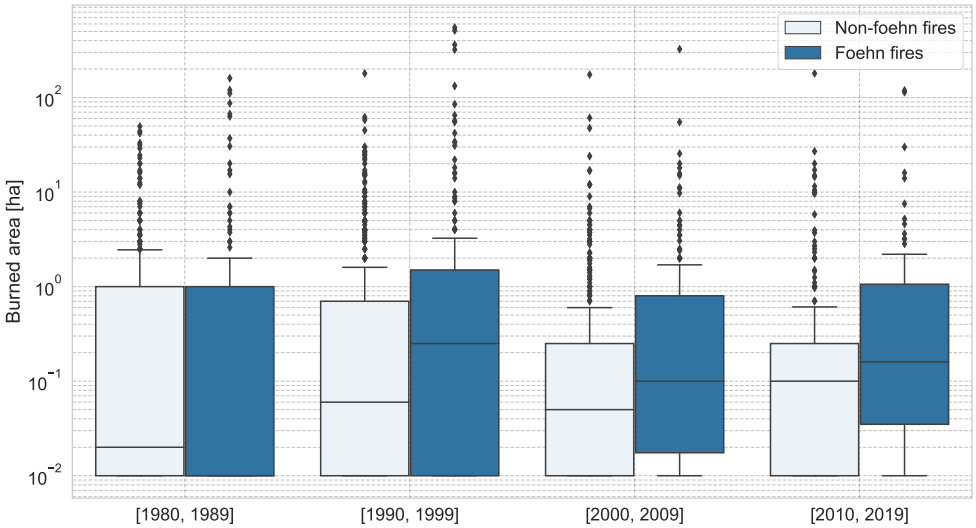


Figure 5.30: Same as Figure 5.28 but disaggregated by decade.

nificance for each foehn versus non-foehn fire plot for the six hour period after fire ignition.

Table 5.13: Shown is the median burned area increase between fires not linked (non-foehn fires) and linked (foehn fires) to foehn in the six hours after ignition. For each result the corresponding statistical significance is shown as determined by a Wilcoxon rank sum test with $\alpha = 0.05$ and Bonferroni correction within each category.

	median increase	p-value	Significant?
All fires	3.0	$5.2 \cdot 10^{-8}$	Yes
Fire regime			
Winter anthropogenic	4.1	$2.4 \cdot 10^{-5}$	Yes
Summer anthropogenic	1.7	$3.5 \cdot 10^{-1}$	No
Summer natural	5.0	$4.0 \cdot 10^{-1}$	No
Foehn location			
North foehn	3.2	$1.2 \cdot 10^{-5}$	Yes
South foehn	2.0	$2.7 \cdot 10^{-3}$	Yes
Decade			
[1980, 1989]	0.5	$7.5 \cdot 10^{-1}$	No
[1990, 1999]	4.2	$4.1 \cdot 10^{-5}$	Yes
[2000, 2009]	2.0	$4.0 \cdot 10^{-4}$	Yes
[2010, 2019]	1.6	$6.6 \cdot 10^{-3}$	Yes

Lastly, we investigated the foehn strength impact on the burned area. We divided wind speed FF and wind speed gusts FFX into three categories, ordered by their value: low, medium, and high. The plots and concrete

values for each bin are shown in Figure 5.31 and Figure 5.32. The tests for statistical significance can be found in Table 5.14.

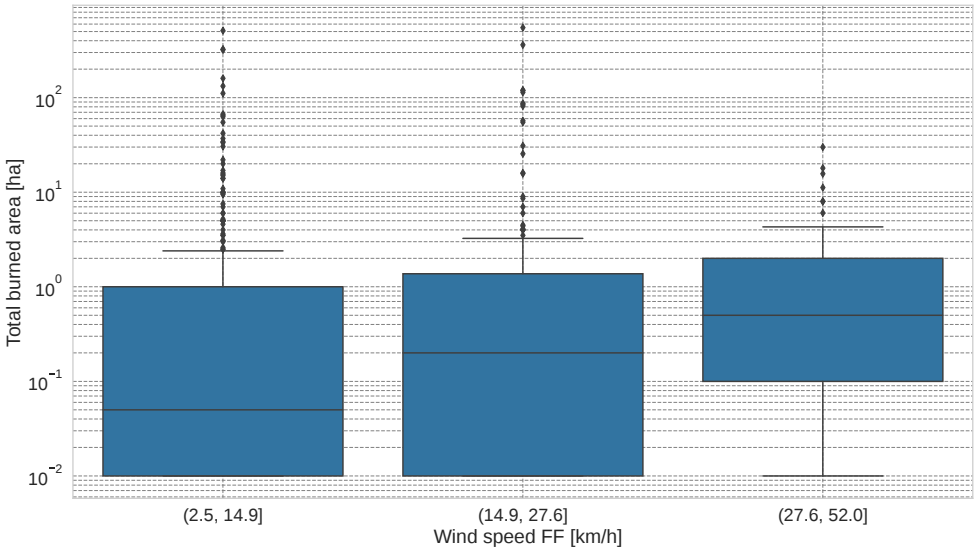


Figure 5.31: Burned area distribution over foehn wind speed FF in the six hours after ignition.

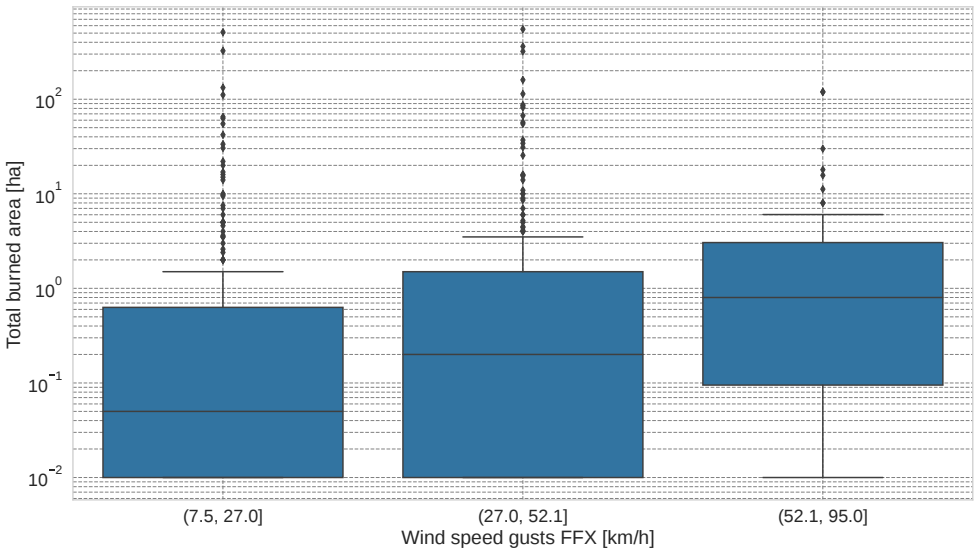


Figure 5.32: Burned area distribution over foehn wind speed gusts FFX in the six hours after ignition.

All plots and tables shown here can also be found for the twelve-hour period in Appendix C.

Table 5.14: Same as Table 5.13 but for FF and FFX.

	median increase	p-value	Significant?
Wind speed FF			
low-medium	4.0	$6.8 \cdot 10^{-3}$	Yes
low-high	10.0	$5.1 \cdot 10^{-3}$	Yes
medium-high	2.5	$1.4 \cdot 10^{-1}$	No
Wind speed gusts FFX			
low-medium	4.0	$1.4 \cdot 10^{-3}$	Yes
low-high	16.0	$8.5 \cdot 10^{-5}$	Yes
medium-high	4.0	$2.3 \cdot 10^{-2}$	No

5.2.2 Discussion

Prior fire ignition

In Figure 5.23, we observed a monotonically increasing relationship between the normalized count of fires and the observed amount of foehn minutes for the 24-hour period. Fires appeared approximately three to seven times more often on days that showed at least four hours of foehn. Also, in the direct comparison between non-foehn and foehn situations in Figure 5.27, fires occurred 3.4 times more often on days where foehn was present. Similar values were obtained for the 48-hour period in Appendix C.

The former paragraph is further underlined by Figure 5.25, where we observed a similar increasing relationship with the temperature increase of a foehn wind. Here the effect was especially pronounced for higher foehn temperature increases and ranged from 1.1 to 1.8. The 48-hour period in Appendix C showed an even more striking increase ranging from 1.1 to 2.5.

Both findings coincide well with the statements from Pezzatti et al. (2016), Audelan (2018), and Sharples (2018) that foehn quickly dries out forests and thus hugely increases the risk of ignition. Hence, we conclude the first objective with the answer that the number of observed fires grows substantially with both the foehn duration and foehn temperature increase.

Past fire ignition

Next, we investigated the impact of foehn on the burned area for the period after fire ignition. Again, in Figure 5.26, we observed an increase in the burned area the longer foehn was prevalent during the six hours after fire start. However, it appeared that a minimum threshold of foehn presence was required to increase the burned area substantially. From the diagram, this turned out to be between one to two hours. If measuring the effect of the foehn duration on the burned area, we observed a two- to sevenfold increase in median between the distribution of fires with no foehn occurrence and the distribution of fires with two or more hours of foehn occurrence.

By comparing non-foehn fires and foehn fires side-by-side in Figure 5.27, we obtained a highly significant median increase of a factor of three (see Table 5.13). These results transfer well when compared to the figures and tables for the twelve-hour period in Appendix C. Here, we estimated a two- to tenfold median increase for the binned diagram and a highly significant twofold median increase for the contrast plot. These results agree well with Sharples (2018), who also found a sixfold increase in the Forest Fire Danger Index (FFDI) for an eight-hour-long foehn event. A sixfold increase in the FFDI implies a concurrent increase in fire behavior characteristics like rate of spread or flame height (Sharples, 2018).

In Figure 5.28, we plotted the results for the different fire regimes. The winter anthropogenic regime (61.6% of all fires) stood out with a statistically significant 4.1 times higher burned area for foehn fires. Summer anthropogenic (25.6% of all fires) and summer natural (12.8% of all fires) did show an increase, however, this turned out to be not significant (see Table 5.13). The twelve-hour time window showed approximately the same results, only here we observed a significant median burned area increase of 3.3 for the winter anthropogenic regime. Again, the other regimes did not show any significant increase. Our findings match with Pezzatti et al. (2016), who also argued that north foehn has its greatest impact in the vegetation rest period when it quickly dries out the litter layer of chestnut forests.

When looking at the different foehn locations in Figure 5.29, we observed a median increase of 3.2 for north foehn fires (77.4% of all fires). For south foehn fires (22.6% of all fires), we measured an increase of two. Again, for both cases, we observed these results to be statistically significant. However, investigating whether the distribution of north foehn fires is larger than the distribution of south foehn fires was highly insignificant (p-value of 0.53). Thus, one cannot conclude that north foehn is more severe than south foehn, and the median deviance could result from statistical noise. An investigation of the twelve-hour time window yielded the same results (see Table C.1). When comparing to the literature, Cetti et al. (2015) stated that the increase in temperature is higher for south foehn than for north foehn due to more poleward origin of north foehn air. If this were the case, one would expect the south foehn to be more potent after adequately controlling for environmental variables (e.g., difference in vegetation between southern and northern Alps). However, no more sources are given in Cetti et al. (2015), and thus we could not verify this statement.

In Figure 5.30, we checked whether there was a trend discernible over the last decades. First, during the 1980 until 1989 period, we observed a non-significant decrease between foehn and non-foehn fires. This is very likely due to the fact that back then, the burned area was not documented well and thus got imputed by 0.01 hectares far more often than the other periods

(see Section 3.2). Second, foehn data started to get available at most stations between 1983 and 1987. Thus, the little existence of foehn data could also have affected this result. In the other decades, we calculated a significant median increase by 4.2, 2.0, and 1.5. Similar values can be observed for the twelve-hour period (see Table C.1), only the last decade did not show a median increase. However, as we obtain from Figure 5.30, it might be useful not only to consider the median but the distribution as a whole. The upper and lower quartile mostly proved to be higher for foehn fires compared to non-foehn fires. Furthermore, when comparing the distributions of foehn fires for the last three decades, no statistically significant decrease can be conducted. Hence, putting everything together, we can conclude that in the last three decades, foehn fires have been significantly more severe than non-foehn fires. However, we cannot conclude that foehn fires became less severe over the same time. Also, De Angelis et al. (2015) stated that homogeneous fire conditions regarding the burned area should have existed since 1978 due to a fire brigade reorganization. This would underline our finding that we did not find any significant change in the burned area of foehn fires over the decades.

In Figure 5.31 and Figure 5.32, we controlled for different foehn strengths during the ignition period. First, we found a statistically significant increase in the burned area between the low-medium and low-high categories for both FF and FFX. If measured by the median, we found this increase to be four and ten, or four and sixteen, respectively. When using the less restrictive Bonferroni-Holm correction instead of the Bonferroni correction, the medium-high increase for FFX would also be significant. However, for consistency, we decided to stick with the Bonferroni correction here. Again, these results agree well with what is expected from the foehn-fire-interaction mechanism (see Subsection 1.2.2) that stronger wind and wind gusts cause more severe flame spread (Byram, 1959).

Finally, we can also conclude the last objective. We found that foehn winds impacted the burned area of forest fires in the past ignition period significantly through their duration as well as exhibited wind speeds.

Chapter 6

Conclusion & Outlook

6.1 Future development of foehn winds

In this part, we established a projection for the foehn frequency under a warming future climate. For this reason, we identified foehn from coarse reanalysis data, then generalized the algorithms to freely running climate simulations, and tested our results extensively. We conclude by summarizing our main findings related to the objectives stated in Subsection 1.1.3.

1. First, we investigated whether it is possible to infer the synoptic foehn situation from coarse NWP data in the Alpine region. We found that the synoptic situation identified by the model corresponds well with what someone would expect from prior physical knowledge. The features which were identified by the random forest are highly interpretable. However, they have to be treated carefully due to their spatial correlation.

2. Second, we scrutinized with which skill the model could identify foehn from such NWP data. Averaged over a ten-year period, we managed to achieve a *Precision* and *Recall* score of 0.786. This corresponds well to the work of other researchers who obtained similar scores on higher-resolution NWP data. Hence, this suggests that the information contained in a coarse reanalysis suffices to predict foehn.

3. Third, we saw that with sufficient feature preprocessing to compensate for biases (potentially due to a different model topography), the transition from a reanalysis to a freely-running climate simulation is feasible. The identified synoptic fingerprint agrees well between reanalysis and freely running climate simulations. For south foehn, the synoptic situation could be determined from a few, well-selected parameters. When comparing the skill of the monthly models, we found the highest skill during spring, autumn, and winter with a $F1$-score ranging from 0.71

to 0.85. During the summer months, this skill degraded to approximately 0.65, where a possible explanation can be found in local-scale mechanisms that counteract the archetypal synoptic fingerprint.

4. Fourth, we investigated the transition from a present-day climate to a warming future climate. Besides a considerable ensemble spread, we found a significant increase in the foehn frequency between CESM-p and CESM-f during February, March, and April, and thus expect foehn to become more common during those months. For May and October, we obtained a significant decrease, and consequently anticipate foehn to become less frequent.

5. Fifth, we employed the same methodology for north foehn, here achieving a comparable skill. We obtained a significant increase between CESM-p and CESM-f for May and July. Moreover, we observed a significant decrease in January, August, and September. For this reason, we reckon foehn to become more or less common during those months, respectively.

During our work, we applied many preprocessing steps in order to make a viable transition from ERAI to CESM. The reason was the different topography and grid between ERAI and CESM. In the future, applying our methodology to a reanalysis and a climate simulation, which have a more similar (potentially even the same) topography and grid, would decrease the number of necessary assumptions to obtain a prediction.

Finally, the methodology we have described in this work is, in principle, applicable also to other regions in the world (e.g., the Rocky Mountains) since ERAI and CESM are global models. One simply has to specify the area of interest and provide observational foehn data from a measurement station. The described procedure will automatically and objectively select the most important features and apply this knowledge to make a projection for future climate. In addition, one can learn from the model which synoptic features matter most for foehn in the specified area.

6.2 Impact of foehn on forest fires

In this part, we investigated which influence foehn winds impose on forest fires from a climatological perspective. Thus, we spatially and temporally linked past foehn and fire occurrence with each other, again testing the results rigorously for their significance. Here, we summarize the findings to the objectives from Subsection 1.2.3.

1. First, we explored the period before a potential fire ignition. We found that if the last 24 hours showed at least four foehn hours, these situations are associated with three to seven times more outbreaks com-

pared to days with less than four foehn hours. Contrasting foehn and non-foehn situations directly, we observed a 3.4-fold increase in the number of fires for days that exhibit foehn. Furthermore, we observed a 1.1- to 1.8-fold increase in the number of fires for foehn winds, which exhibit a stronger temperature increase. The 48-hour period showed a similar, even slightly more pronounced, pattern.

2. Second, we investigated the influence of foehn on the fire severity after the start of a fire. We found that depending on the foehn length, the median burned area increased two to sevenfold if more than two foehn hours were present. When comparing foehn and non-foehn fires, we observed a statistically significant increase of a factor of three.

   Additionally, by comparing different fire regimes, we found that this increase stemmed primarily from the winter anthropogenic regime, where foehn increased the median burned area by a significant factor of 4.1. Second, south and north foehn both increased the burned area significantly. However, only an insignificant difference regarding the severity could be observed between south and north foehn fires. When comparing the trend over the decades, we observed that in each of the last three decades, foehn fires have been significantly more severe than non-foehn fires. However, we cannot conclude that foehn fires changed in their severity over the decades.

   Lastly, we also assessed the link between foehn strength, measured by the wind speed FF and wind speed gusts FFX. Again, we found a significant increase in the burned area between different foehn strengths by a factor of four from low to medium foehn strength.

To conclude, we claim to have established a link between foehn and its impact on forest fires in a quantitative manner. On the one hand, foehn severely improves the ignition conditions leading to an increased number of fires. On the other hand, during a fire, it also drives the fire and magnifies the burned area.

One potential enhancement of our work would be to incorporate more control variables into the analysis. Here, in our view, Bayesian or Causal Networks offer an elegant way of introducing variables like soil moisture or available burnable material. For example, we expect soil moisture as a mediator between foehn and the number of fires. Furthermore, the soil moisture will also be impacted by other factors. The structure of such a Causal Network should be constructible from expert knowledge, while the exact relationships between variables can be obtained from data.

In the future, we hope that the results of this work can help to assess the fire risk associated with specific foehn situations. Especially foehn situations, which exhibit a long duration, a large temperature increase, or high wind

speeds severely impact forest fires. In addition, in the first part, we found that foehn will become more common during the already nowadays fire-prone spring and summer season for south and north foehn, respectively. Consequently, we argue once more that an accurate foehn forecast is crucial. Foehn unfolds its impact within a matter of hours, and immediate action is essential to mitigate a potentially large number of fires or burned areas.

Appendix A

Calculation of derived variables

Here, we quickly outline how the derived variables have been calculated. The CESM grid included the latitudes

$$lat \in \{42.87, 43.82, 44.76, 45.70, 46.64, 47.59, 48.53, 49.47\} = M_{lat}$$

(in degrees North) and the longitudes

$$lon \in \{0, 1.25, 2.5, 3.75, 5, 6.25, 7.5, 8.75, 10, 11.25, 12.5, 13.75, 15\} = M_{lon}$$

(in degrees East).

The sea level pressure differences were calculated via

$$\Delta SLP = \Delta SLP_{i,j,k,m} = SLP_{i,j} - SLP_{k,m}$$
$$\text{for } i,k \in M_{lat} \quad j,m \in M_{lon} \quad k > i \quad j > m$$

The geopotential height differences were calculated in the same manner for the pressure levels 850 hPa and 700 hPa

$$\Delta Z = \Delta Z_{i,j,k,m,h} = Z_{i,j,h} - Z_{k,m,h}$$
$$\text{for } i,k \in M_{lat} \quad j,m \in M_{lon} \quad h \in \{850, 700\} \quad k > i \quad j > m$$

The potential temperature was calculated via

$$\theta = T \cdot \left(\frac{p_0}{p}\right)^{\kappa}$$

where $p_0 = 1000$ hPa defines the reference pressure and $\kappa = 0.286$ is the Poisson constant for dry air. The potential temperature simply describes the temperature an air parcel would have if it would be brought to sea level. Thus, it allows to adjust for the varying pressure at different heights.

From the potential temperature, we calculated the horizontal potential temperature differences for the pressure level on 850 hPa

$$\Delta\theta_{hor} = \Delta\theta_{i,j,k,m} = \theta_{i,j} - \theta_{k,m}$$
$$\text{for } i,k \in M_{lat} \quad j,m \in M_{lon} \quad k > i \quad j > m$$

Furthermore, we looked at the stability of the atmosphere, which is resembled by the vertical potential temperature difference. In a stable stratified atmosphere, foehn has a harder time reaching the ground. Thus, we examined the vertical potential temperature differences with reference to a ground-near layer of 900 hPa

$$\Delta\theta_{ver} = \Delta\theta_{i,j,h} = \theta_{i,j,h} - \theta_{i,j,900}$$
$$\text{for } i \in M_{lat} \quad j \in M_{lon} \quad h \in \{850, 700\}$$

Appendix B

Composite plots for CESM-p and CESM-f

B.1 South foehn

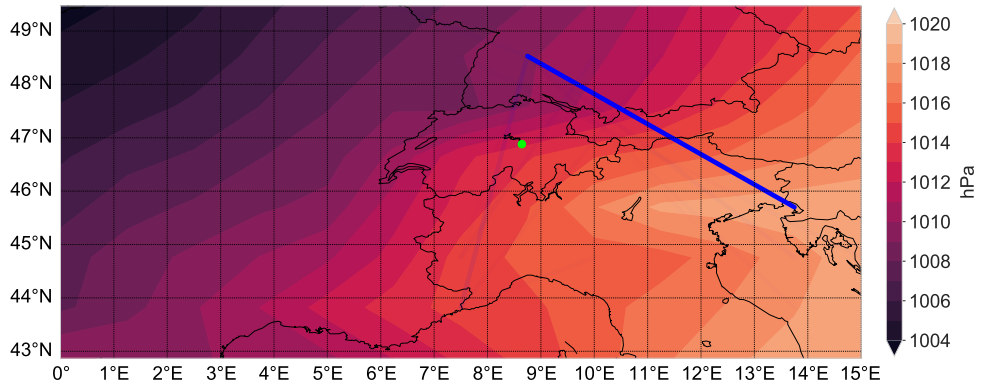


Figure B.1: Shown are the most important sea level pressure differences ΔSLP over the mean SLP weather condition in CESM-p climate for all predicted foehn events. The more opaque the blue lines, the more important are the specific differences. Due to the large feature importance of one feature, the other ones are barely visible. We kept this scaling to underline the fact that one feature appears to be enough to identify south foehn to the largest extent. The green dot marks the location of Altdorf.

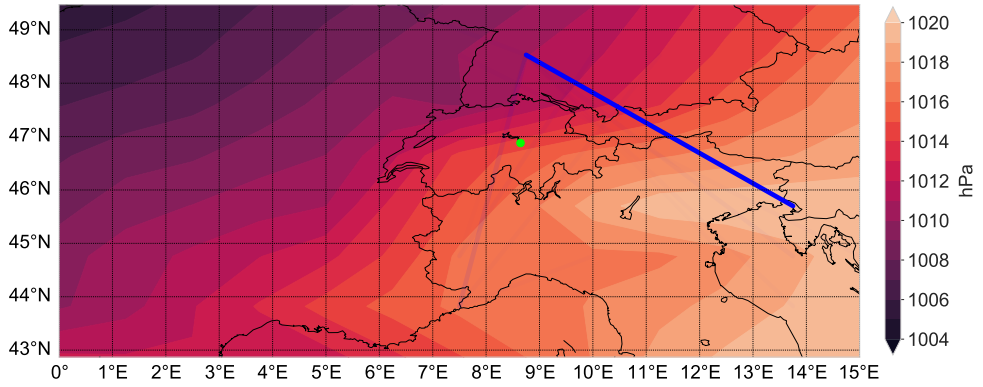


Figure B.2: Same as Figure B.1 but for CESM-f.

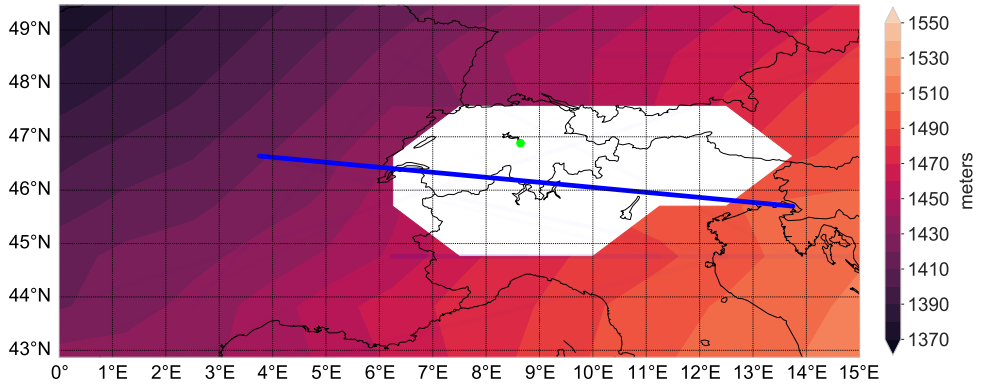


Figure B.3: Same as Figure B.1 but for ΔZ on 850 hPa.

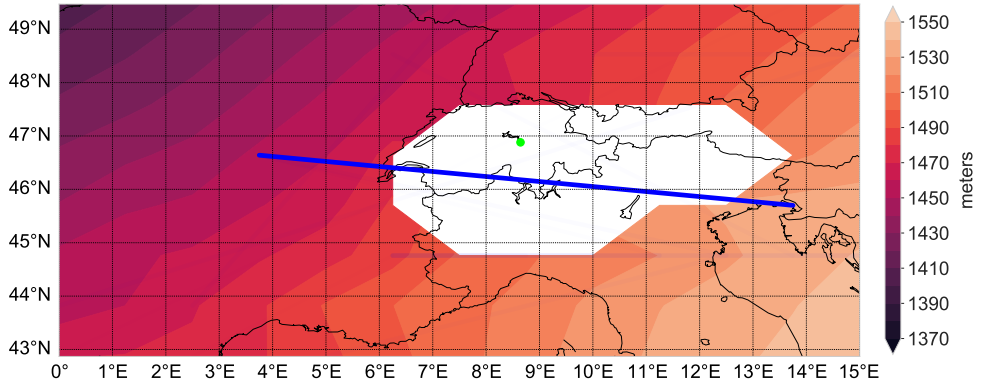


Figure B.4: Same as Figure B.3 but for CESM-f. Note how the 850 hPa pressure level appears to be shifted upward by about 40 meters compared to present climate.

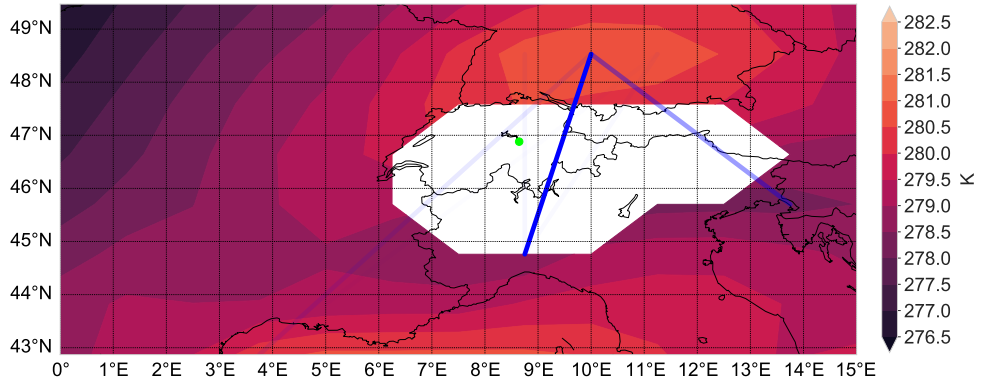


Figure B.5: Same as Figure B.1 but for $\Delta\theta_{hor}$ on 850 hPa.

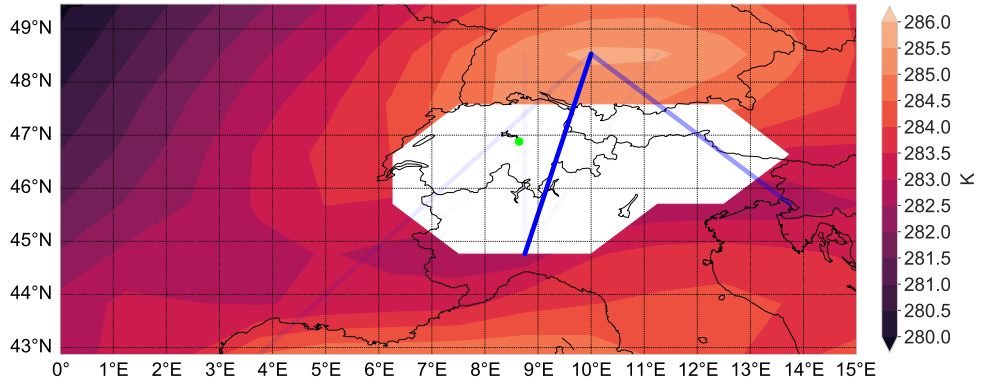


Figure B.6: Same as Figure B.5 but for CESM-f. Note that we needed to adjust the scale by shifting it up by 3.5 K compared to the composite plots for ERAI and CESM-p to retain readability.

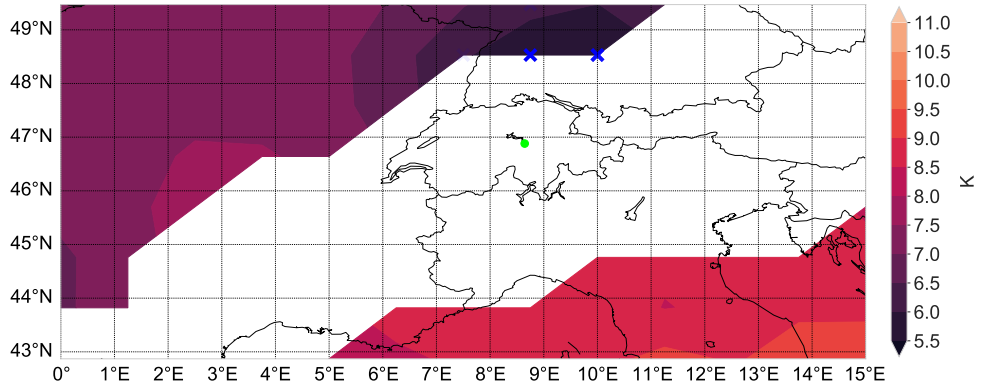


Figure B.7: Same as Figure B.1 but for $\Delta\theta_{ver}$ from 900 to 700 hPa. The more opaque the blue cross, the more important is the specific difference.

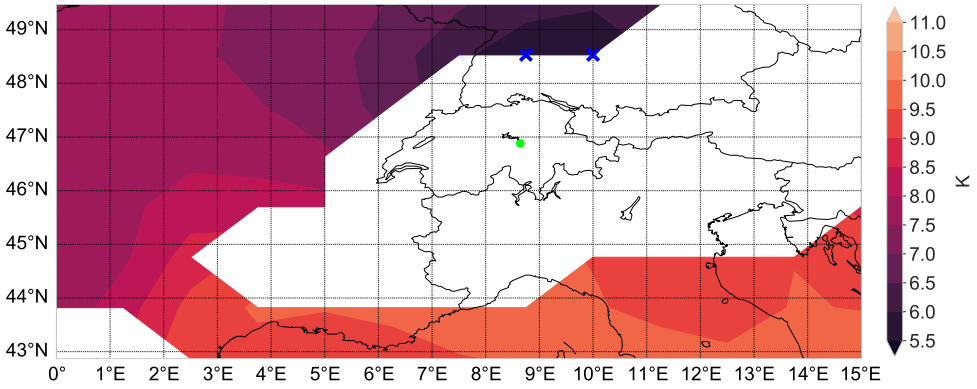


Figure B.8: Same as Figure B.7 but for CESM-f.

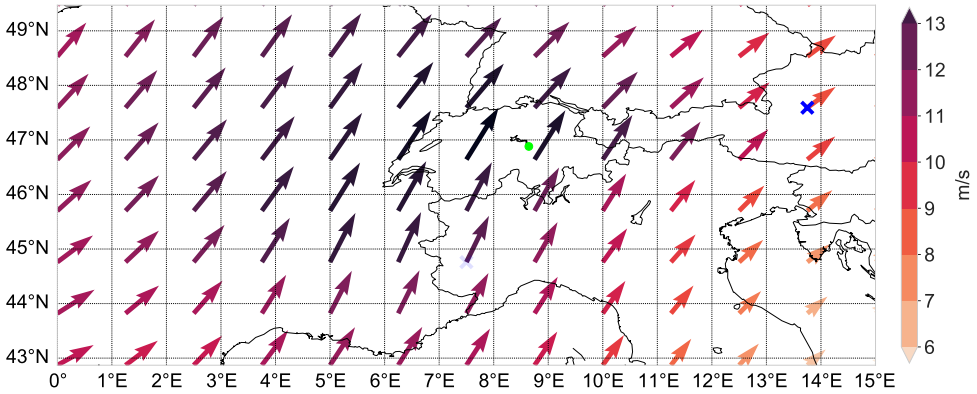


Figure B.9: Same as Figure B.7 but for U and V on 700 hPa.

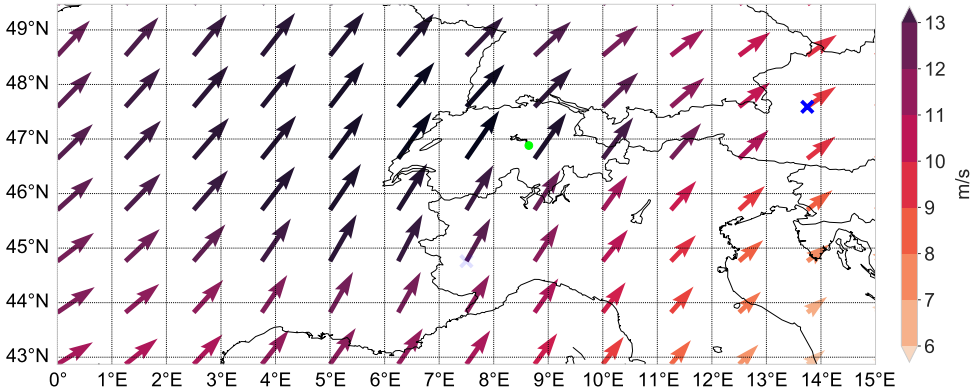


Figure B.10: Same as Figure B.9 but for CESM-f.

B.2 North foehn

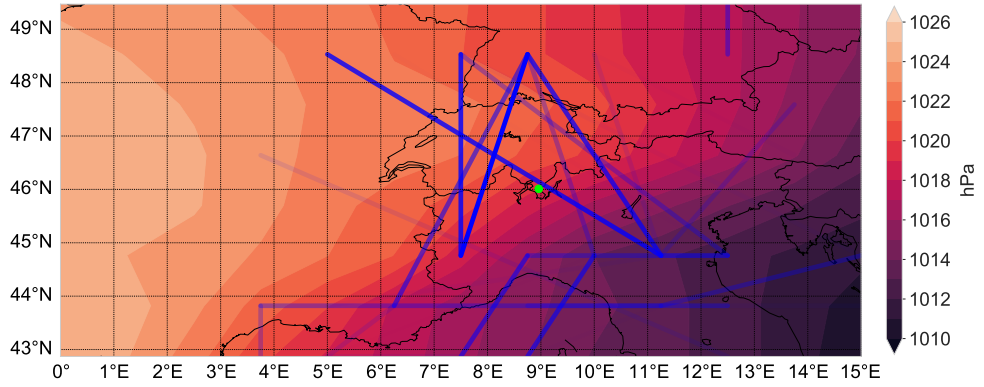


Figure B.11: Same as Figure B.1 but for north foehn. Note how now more features appear to have similar importance. The green dot marks the location of Lugano.

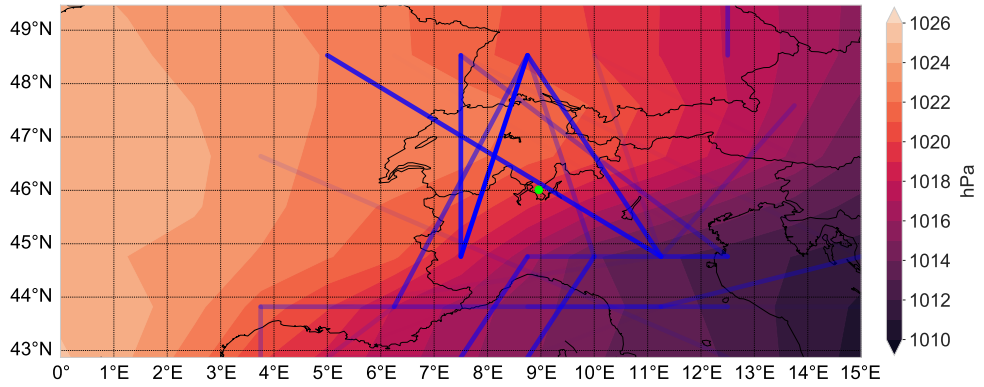


Figure B.12: Same as Figure B.11 but for CESM-f.

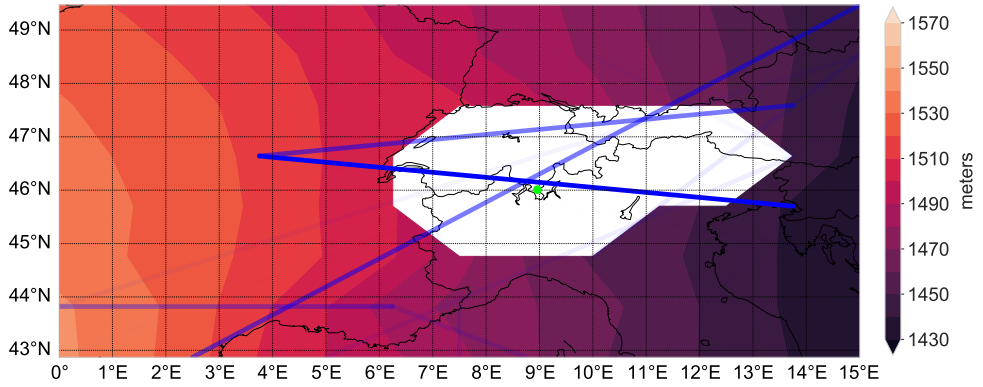


Figure B.13: Same as Figure B.11 but for ΔZ on 850 hPa.

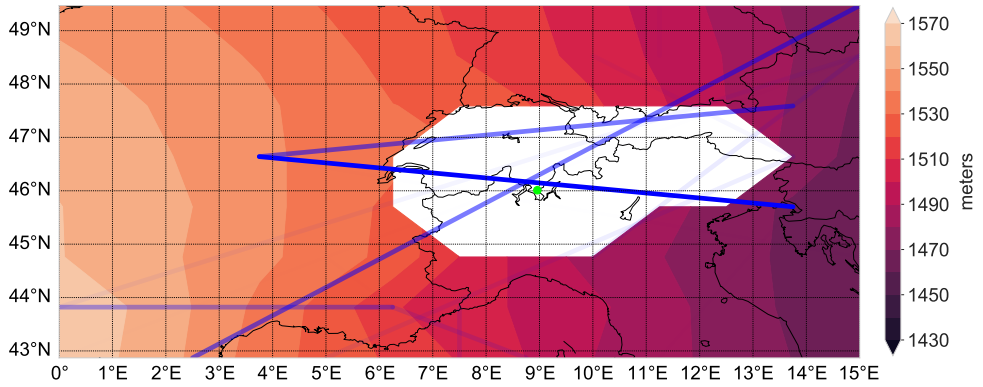


Figure B.14: Same as Figure B.13 but for CESM-f. Note how again the 850 hPa pressure level appears to be shifted upward by about 40 meters compared to present climate.

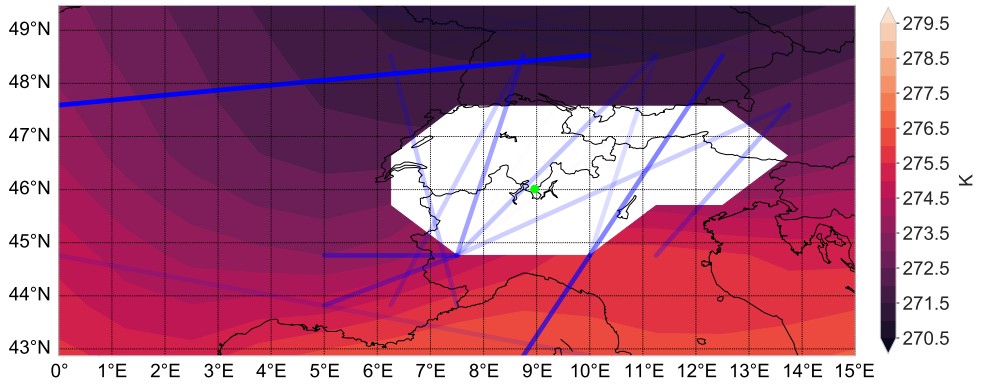


Figure B.15: Same as Figure B.11 but for $\Delta\theta_{hor}$ on 850 hPa.

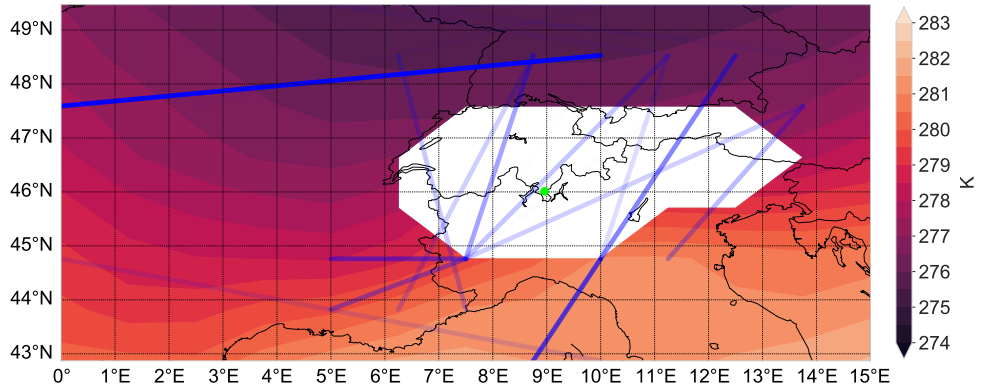


Figure B.16: Same as Figure B.15 but for CESM-f. Note that we needed to adjust the scale by shifting it up by 3.5 K compared to the composite plots of ERAI and CESM-p to retain readability.

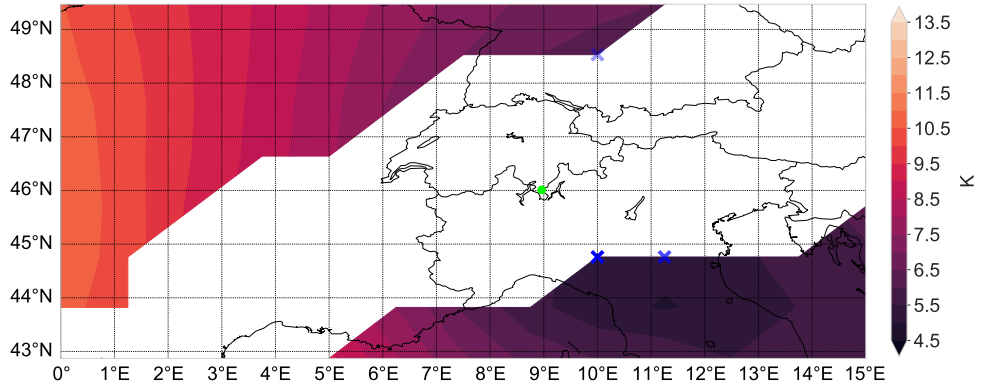


Figure B.17: Same as Figure B.7 but for north foehn. The green dot marks the location of Lugano.

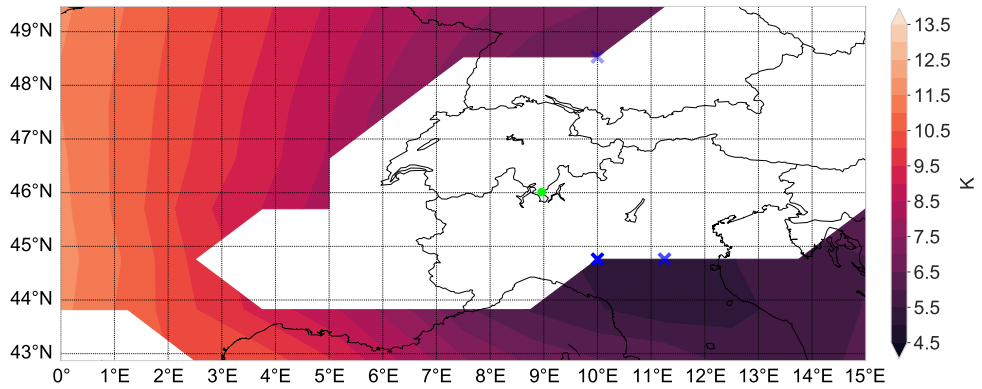


Figure B.18: Same as Figure B.17 but for CESM-f.

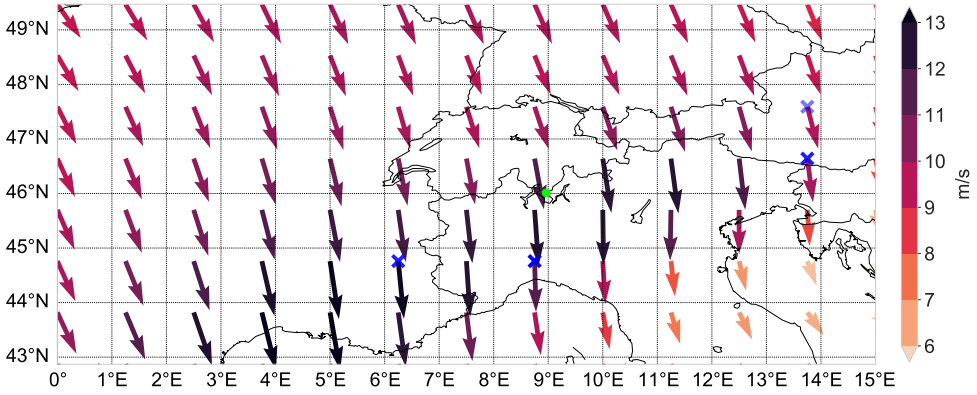


Figure B.19: Same as Figure B.17 but for U and V on 700 hPa.

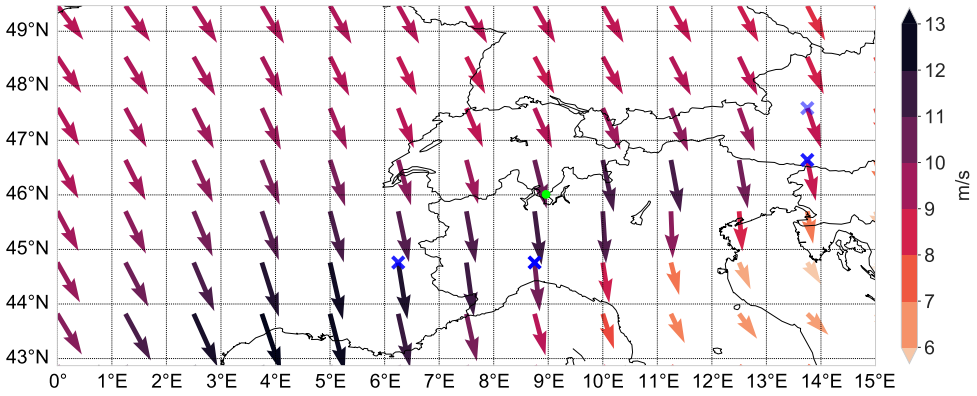


Figure B.20: Same as Figure B.19 but for CESM-f.

Appendix C

Impact of foehn on forest fires

C.1 Prior fire ignition

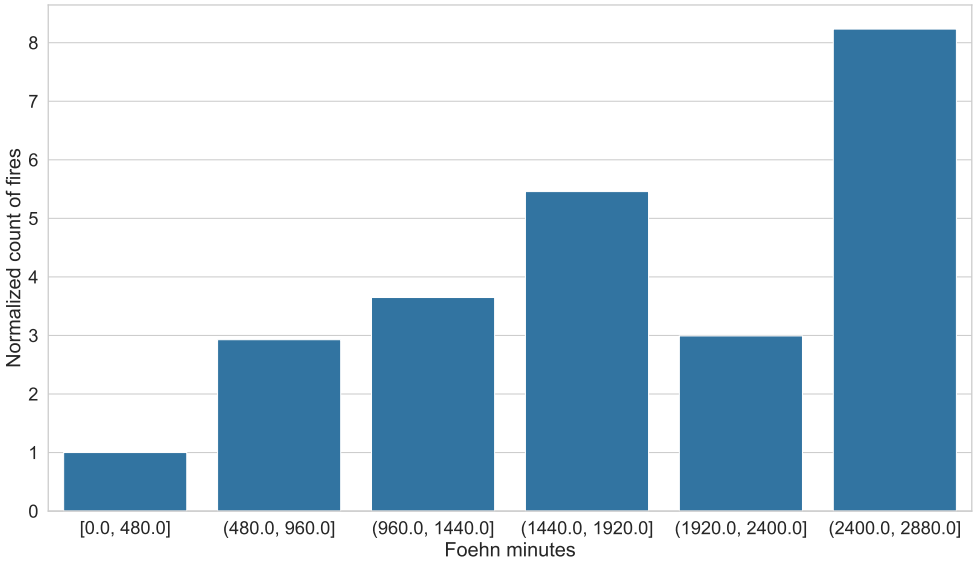


Figure C.1: Same as Figure 5.23 but for the 48-hour period.

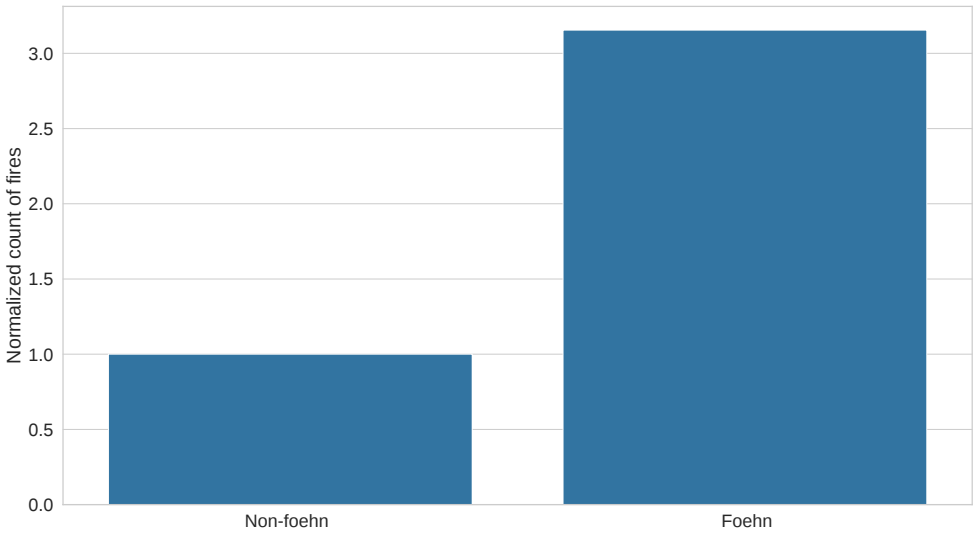


Figure C.2: Same as Figure 5.24 but for the 48-hour period.

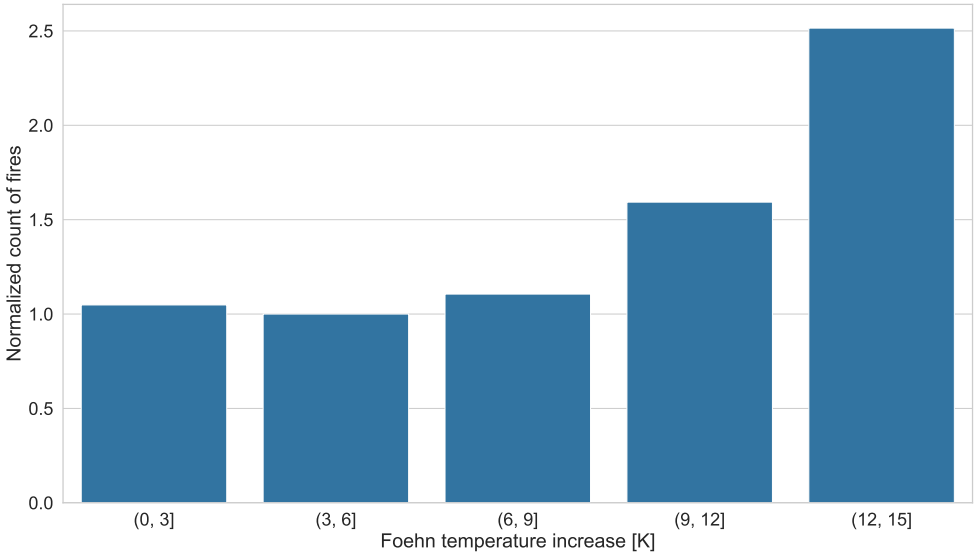


Figure C.3: Same as Figure 5.25 but for the 48-hour period.

C.2 Post fire ignition

C.2.1 Six-hour period

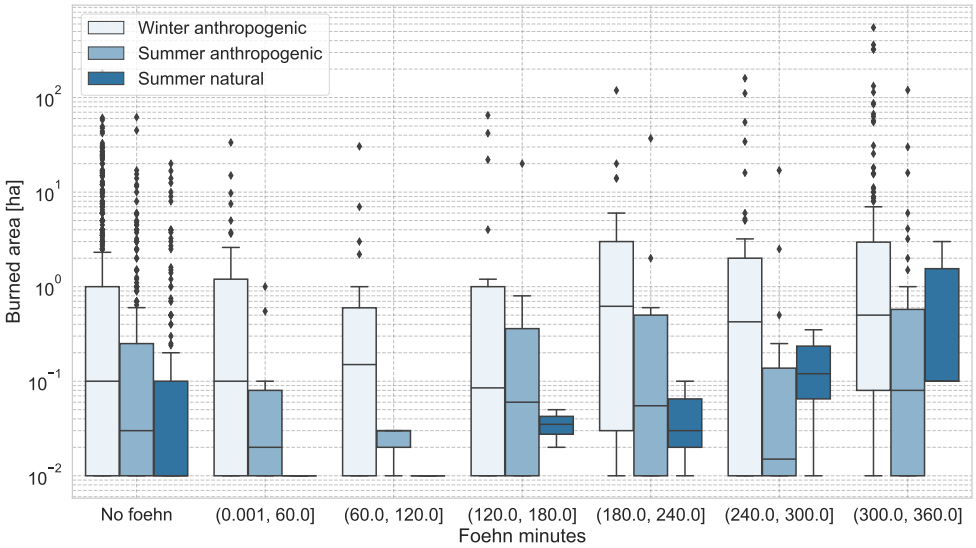


Figure C.4: Same as Figure 5.26 but disaggregated by fire regime.

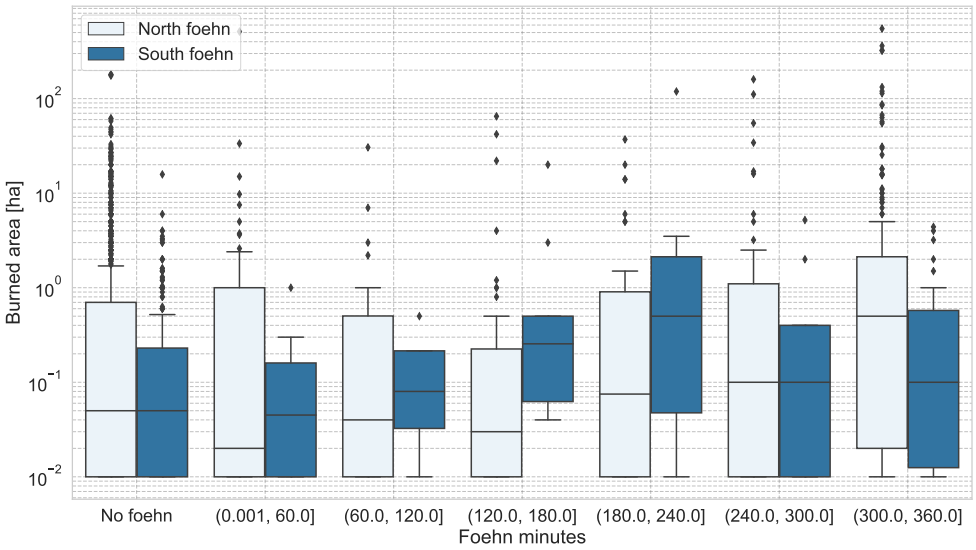


Figure C.5: Same as Figure 5.26 but disaggregated by foehn location.

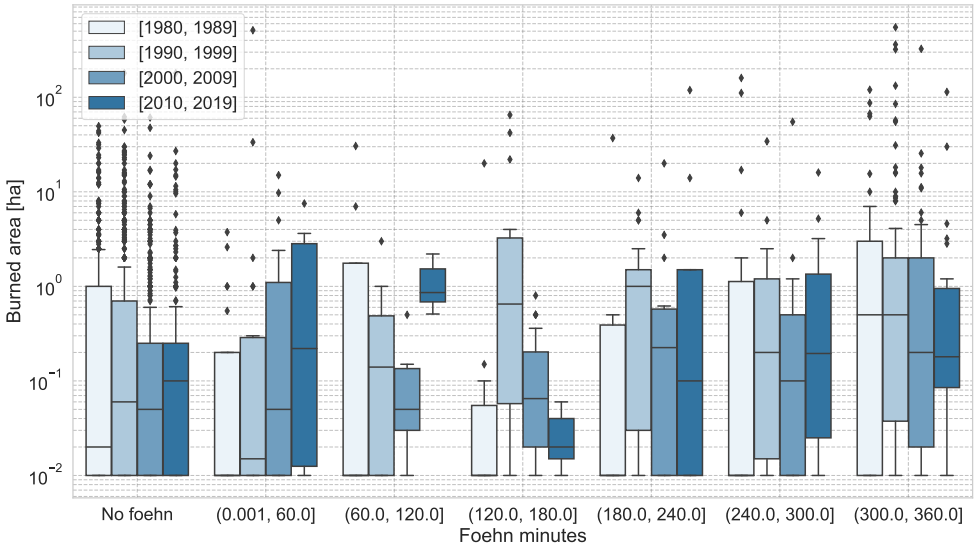


Figure C.6: Same as Figure 5.26 but disaggregated by decade.

C.2.2 Twelve-hour period

Table C.1: Same as Table 5.13 but for the twelve hours after ignition.

	median increase	p-value	Significant?
All fires	2.0	$5.4 \cdot 10^{-7}$	Yes
Fire regime			
Winter anthropogenic	3.3	$1.7 \cdot 10^{-4}$	Yes
Summer anthropogenic	1.7	$2.4 \cdot 10^{-1}$	No
Summer natural	5.0	$4.3 \cdot 10^{-1}$	No
Foehn location			
North Foehn	2.2	$8.8 \cdot 10^{-5}$	Yes
South Foehn	2.0	$2.3 \cdot 10^{-3}$	Yes
Decade			
[1980, 1989]	0.5	$7.8 \cdot 10^{-1}$	No
[1990, 1999]	3.6	$1.9 \cdot 10^{-4}$	Yes
[2000, 2009]	2.0	$5.4 \cdot 10^{-4}$	Yes
[2010, 2019]	1.0	$1.6 \cdot 10^{-2}$	Yes

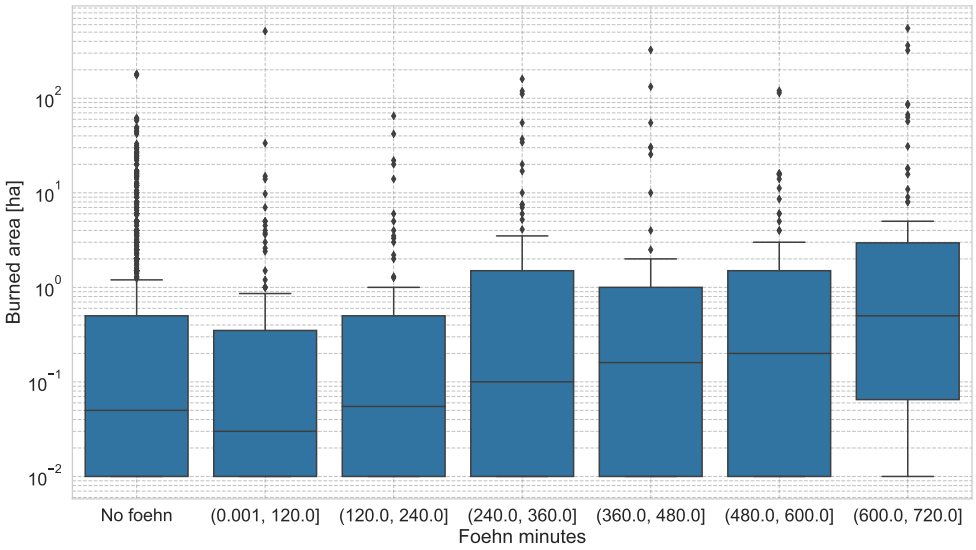


Figure C.7: Same as Figure 5.26 but for the twelve hours after ignition.

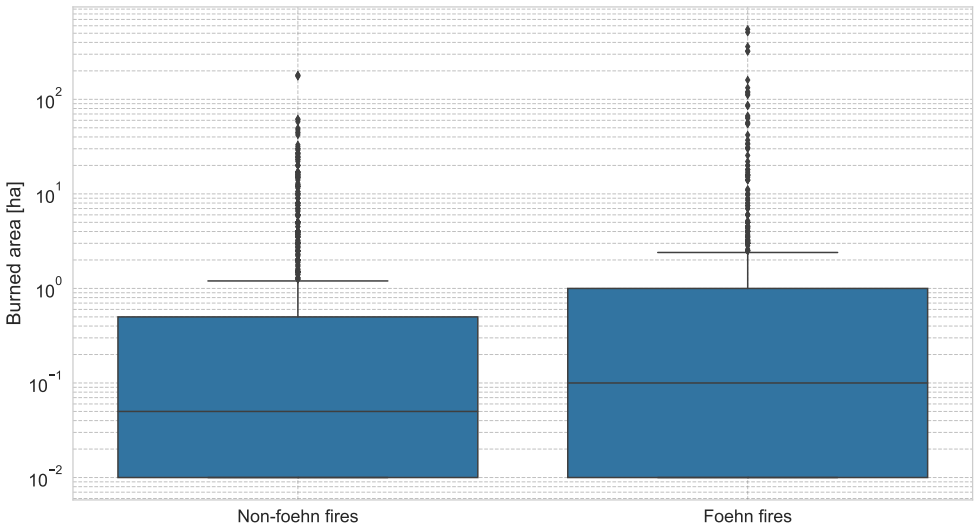


Figure C.8: Same as Figure 5.27 but for the twelve hours after ignition.

Fire regime

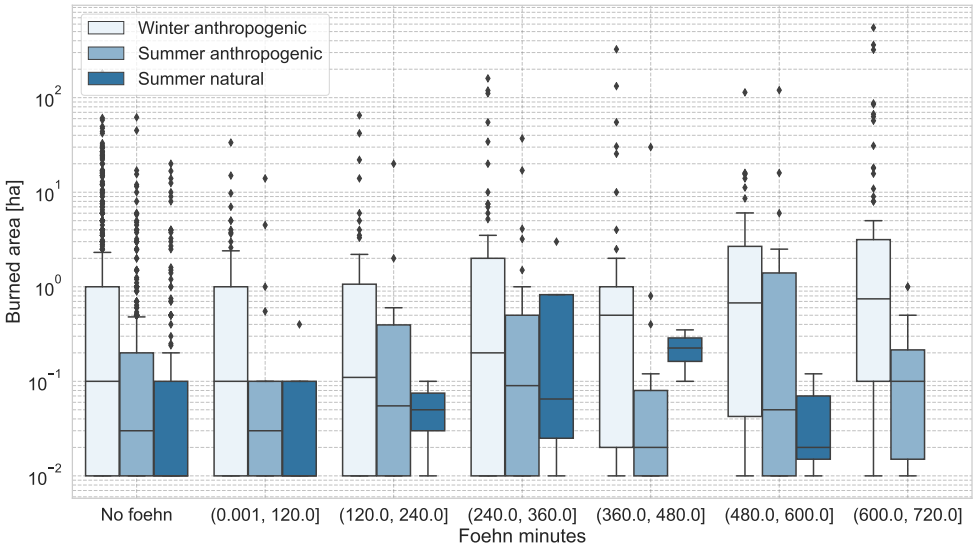


Figure C.9: Same as Figure C.7 but disaggregated by fire regime.

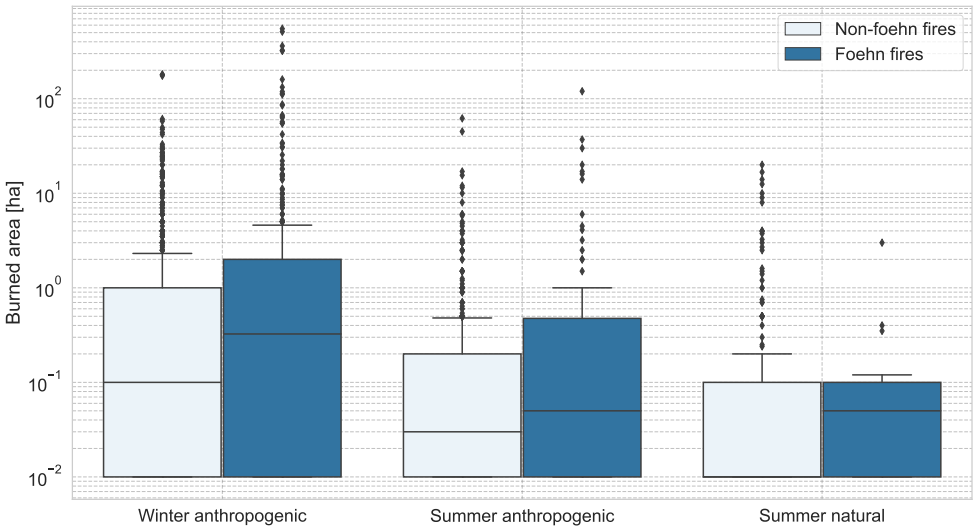


Figure C.10: Same as Figure C.8 but disaggregated by fire regime.

Foehn location

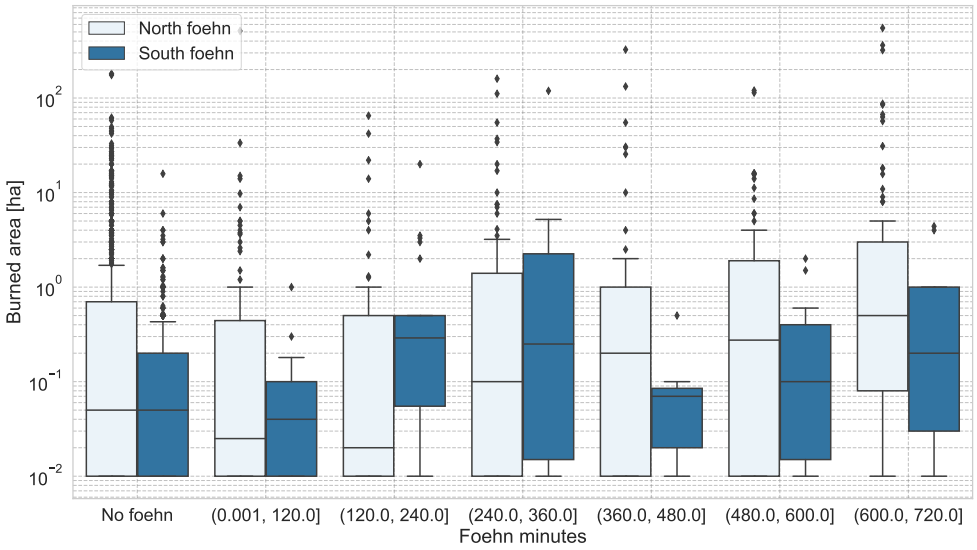


Figure C.11: Same as Figure C.7 but disaggregated by foehn location.

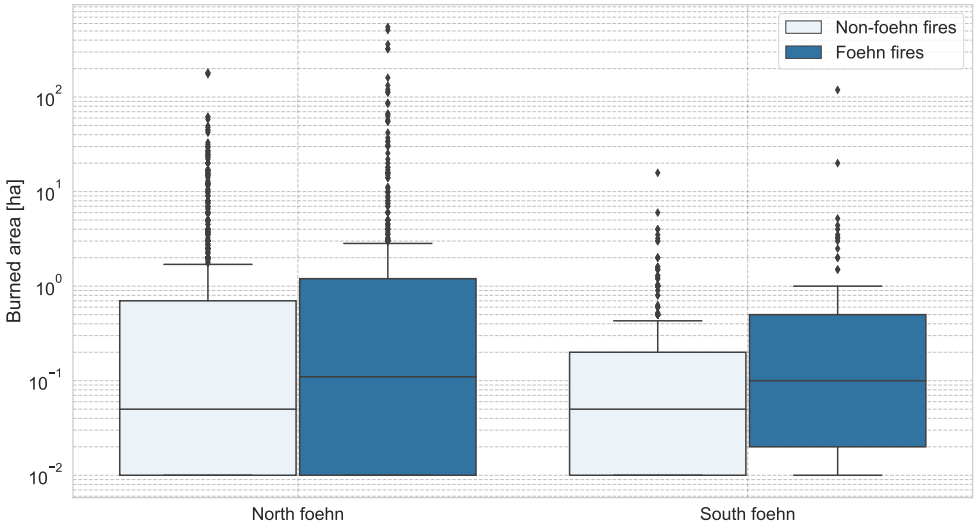


Figure C.12: Same as Figure C.8 but disaggregated by foehn location.

Decade

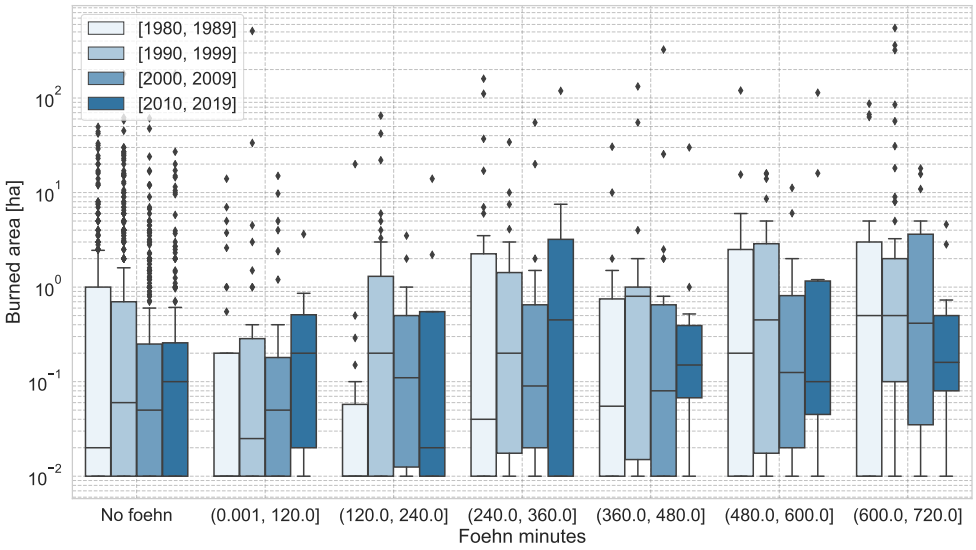


Figure C.13: Same as Figure C.7 but disaggregated by decade.

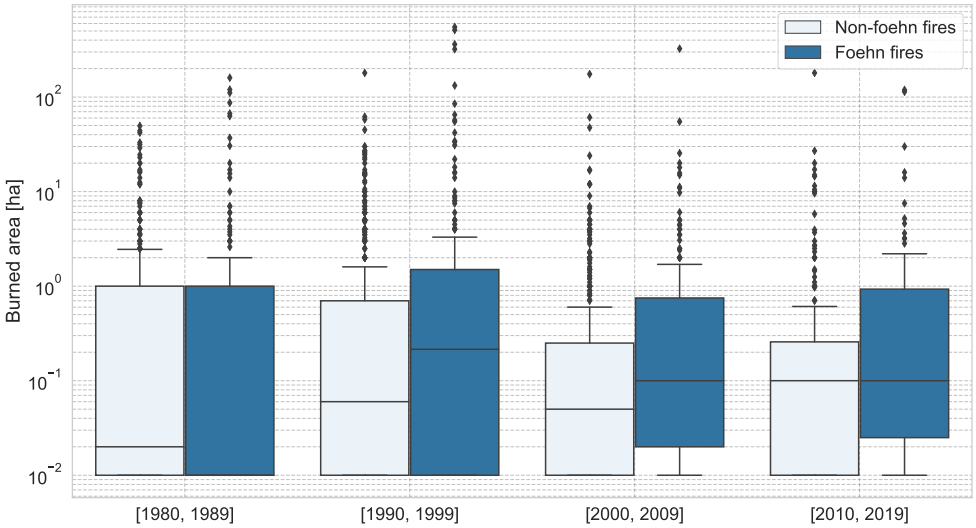


Figure C.14: Same as Figure C.8 but disaggregated by decade.

Foehn strength

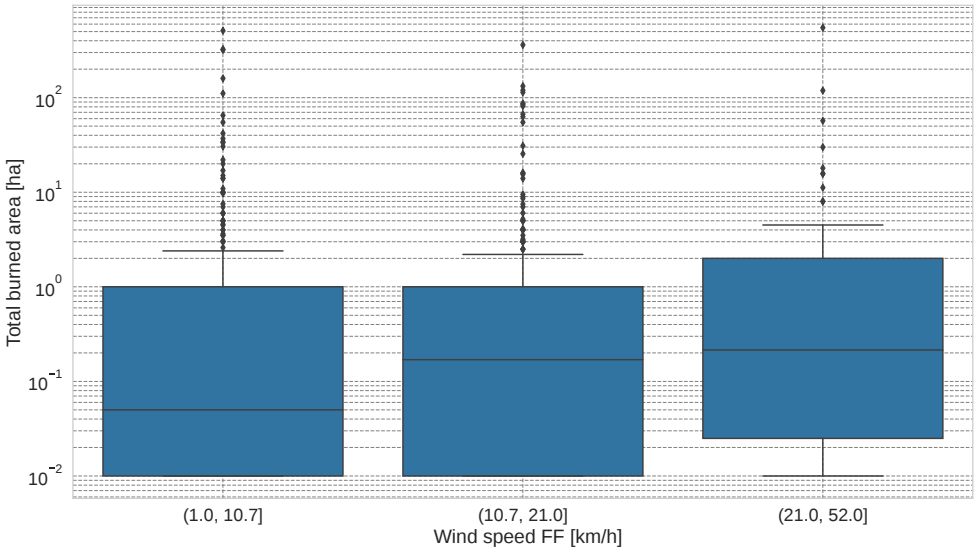


Figure C.15: Same as Figure 5.31 but for the twelve hours after ignition.

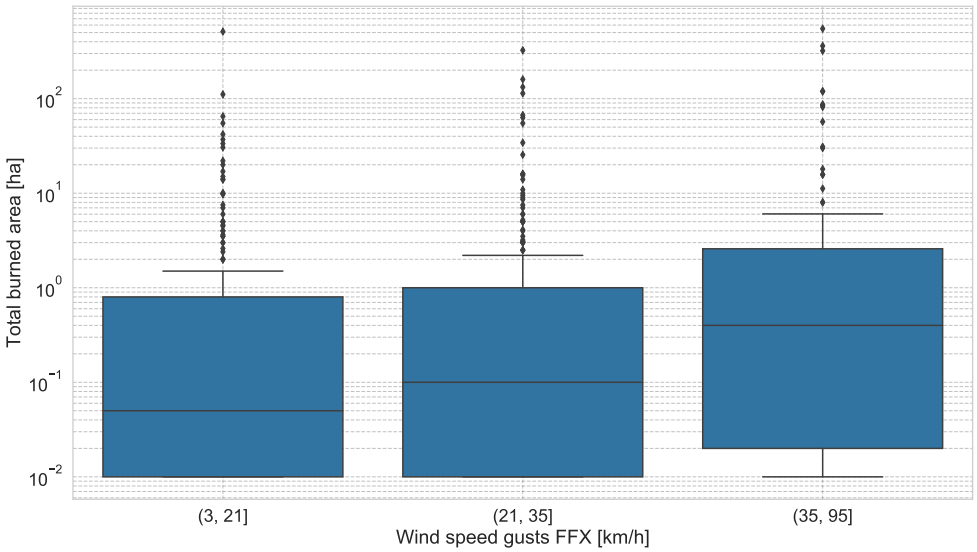


Figure C.16: Same as Figure 5.32 but for the twelve hours after ignition.

Table C.2: Same as Table 5.14 but for the twelve hours after ignition.

	median increase	p-value	Significant?
Wind speed FF			
low-medium	3.4	$2.6 \cdot 10^{-2}$	No
low-high	4.3	$4.6 \cdot 10^{-3}$	Yes
medium-high	1.3	$1.3 \cdot 10^{-1}$	No
Wind speed gusts FFX			
low-medium	2.0	$1.1 \cdot 10^{-1}$	No
low-high	8.0	$2.3 \cdot 10^{-4}$	Yes
medium-high	4.0	$1.1 \cdot 10^{-1}$	Yes

Bibliography

Audelan, A., 2018. Improvement of the fire index used in Ticino, in Switzerland and Application of this method to the Hautes-Alpes in France. Technical report, WSL. 55 pp.

Beusch, L., Raveh-Rubin, S., Sprenger, M., and Papritz, L., 2018. Dynamics of a Puelche foehn event in the Andes. *Meteorologische Zeitschrift*, 27:67–80.

Breiman, L., 2001. Random forests. *Machine learning*, 45:5–32.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A., 1984. *Classification and regression trees*. CRC press. 368 pp.

Burri, K., Dürr, B., Gutermann, T., Neururer, A., Werner, R., and Zala, E., 2007. Foehn Verification with the COSMO Model. In *International Conference on Alpine Meteorology (ICAM)*, Chambéry, France.

Byram, G., 1959. Combustion of forest fuels. In *Forest fire: control and use*, pages 61–89. McGraw-Hill, New York, NY.

Cape, M., Vernet, M., Skvarca, P., Marinsek, S., Scambos, T., and Domack, E., 2015. Foehn winds link climate-driven warming to ice shelf evolution in Antarctica. *Journal of Geophysical Research: Atmospheres*, 120:11–37.

Carrega, P., 1991. A Meteorological Index of Forest Fire Hazard in Mediterranean France. *International Journal of Wildland Fire*, 1:79–86.

Cetti, C., Buzzi, M., and Sprenger, M., 2015. Climatology of Alpine north foehn. *Scientific Report MeteoSwiss*, 100. 76 pp.

Chen, T. and Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd international conference on knowledge discovery and data mining*, pages 785–794.

Chen, T. and He, T., 2015. Higgs boson discovery with boosted trees. In *NIPS 2014 workshop on high-energy physics and machine learning*, pages 69–80.

Conedera, M., Cesti, G., Pezzatti, G., Zumbrunnen, T., and Spinedi, F., 2006. Lightning-induced fires in the Alpine region: An increasing problem. *Forest Ecology and Management*, 234:9 pp.

Conrad, V., 1936. *Die klimatologischen Elemente und ihre Abhängigkeit von terrestrischen Einflüssen*. Schweizerbart Science Publishers. 556 pp.

Courvoisier, H. W. and Gutermann, T., 1971. Zur praktischen Anwendung des Föhntests von Widmer. *Arbeitsberichte der MeteoSchweiz*, 21:7 pp.

Davis, J. and Goadrich, M., 2006. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240.

De Angelis, A., Ricotta, C., Conedera, M., and Pezzatti, G. B., 2015. Modelling the meteorological forest fire niche in heterogeneous pyrologic conditions. *PloS one*, 10:17 pp.

Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., and others,, 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137: 553–597.

Drechsel, S. and Mayr, G. J., 2008. Objective forecasting of foehn winds for a subgrid-scale alpine valley. *Weather and Forecasting*, 23:205–218.

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

Elvidge, A. D. and Renfrew, I. A., 2016. The causes of foehn warming in the lee of mountains. *Bulletin of the American Meteorological Society*, 97:455–466.

Fink, A. H., Brücher, T., Krüger, A., Leckebusch, G. C., Pinto, J. G., and Ulbrich, U., 2004. The 2003 European summer heatwaves and drought–synoptic diagnosis and impacts. *Weather*, 59:209–216.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 25:1189–1232.

Gerstgrasser, D., 2017. Dokumentation Südföhn. Technical report, MeteoSwiss. 59 pp.

Glahn, H. R. and Lowry, D. A., 1972. The use of model output statistics (MOS) in objective weather forecasting. *Journal of applied meteorology*, 11: 1203–1211.

Gutermann, T., 1970. *Vergleichende Untersuchungen zur Föhnhäufigkeit im Rheintal zwischen Chur und Bodensee*. City-Druck AG. 68 pp.

Gutermann, T., Dürr, B., Richner, H., and Bader, S., 2012. Föhnklimatologie Altdorf: die lange Reihe (1864-2008) und ihre Weiterführung, Vergleich mit anderen Stationen. *Fachbericht MeteoSchweiz*, 241:53 pp.

Guzman-Morales, J., Gershunov, A., Theiss, J., Li, H., and Cayan, D., 2016. Santa Ana Winds of Southern California: Their climatology, extremes, and behavior spanning six and a half decades. *Geophysical Research Letters*, 43: 2827–2834.

Hächler, P., Burri, K., Dürr, B., Gutermann, T., Neururer, A., Richner, H., and Werner, R., 2011. Der Föhnfall vom 8. Dezember 2006 - eine Fallstudie. *Arbeitsberichte der MeteoSchweiz*, 234:56 pp.

Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media. 745 pp.

Hoens, T. R. and Chawla, N. V., 2013. Imbalanced datasets: From sampling to classifiers. In *Imbalanced Learning*, pages 43–59. John Wiley & Sons, Ltd.

Johnson, R. and Zhang, T., 2013. Learning nonlinear functions using regularized greedy forest. *IEEE transactions on pattern analysis and machine intelligence*, 36:942–954.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S., Danabasoglu, G., Edwards, J., and others,, 2015. The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96:1333–1349.

Keeley, J. E., 2004. Impact of antecedent climate on fire regimes in coastal California. *International Journal of Wildland Fire*, 13:173–182.

Kljun, N., Sprenger, M., and Schär, C., 2001. Frontal modification and lee cyclogenesis in the Alps: A case study using the ALPEX reanalysis data set. *Meteorology and Atmospheric Physics*, 78:89–105.

Laffin, M., Zender, C., Singh, S., and van Wessem, M., 2019. 40 years of foehn winds on the antarctic peninsula: Impact on surface melt from 1979-2018. In *AGU Fall Meeting*.

Littell, J., Peterson, D., Riley, K., Liu, Y., and Luce, C., 2016. A review of the relationships between drought and forest fire in the United States. *Global change biology*, 22:2353–2369.

Lotteraner, C. J., 2009. *Synoptisch-klimatologische Auswertung von Windfeldern im Alpenraum*. PhD thesis, University Vienna. 112 pp.

Louppe, G., 2014. *Understanding random forests: From theory to practice*. PhD thesis, Université de Liège. 223 pp.

Mcgowan, H. and Sturman, A., 1996. Regional and local scale characteristics of foehn wind events over the South Island of New Zealand. *Meteorology and Atmospheric Physics*, 58:151–164.

Mehta, P., Bukov, M., Wang, C.-H., Day, A. G., Richardson, C., Fisher, C. K., and Schwab, D. J., 2019. A high-bias, low-variance introduction to machine learning for physicists. *Physics reports*, 810:1–124.

Mercer, A. E., Richman, M. B., Bluestein, H. B., and Brown, J. M., 2008. Statistical modeling of downslope windstorms in Boulder, Colorado. *Weather and Forecasting*, 23:1176–1194.

Miltenberger, A. K., Reynolds, S., and Sprenger, M., 2016. Revisiting the latent heating contribution to foehn warming: Lagrangian analysis of two foehn events over the Swiss Alps. *Quarterly Journal of the Royal Meteorological Society*, 142:2194–2204.

Murphy, A. H., 1996. The Finley affair: A signal event in the history of forecast verification. *Weather and Forecasting*, 11:3–20.

Norte, F. A., 2015. Understanding and forecasting Zonda wind (Andean foehn) in Argentina: a review. *Atmospheric and Climate Sciences*, 5:163–193.

Oard, M. J., 1993. A method for predicting Chinook winds east of the Montana Rockies. *Weather and Forecasting*, 8:166–180.

Obenland, E., 1956. Untersuchungen zur Föhnstatistik des Oberallgäus. *Berichte des Deutschen Wetterdienstes*, 23:37 pp.

Ólafsson, H., 2005. The heat source of the foehn. *Hrvatski meteorološki časopis*, 40:542–545.

Orlanski, I., 1975. A Rational Subdivision of Scales for Atmospheric Processes. *Bulletin of the American Meteorological Society*, 56:527–530.

Ortner, A., 2010. *Detaillierte Statistik des Südfoehns in Innsbruck: Einfluss des Südföhns auf die Temperatur*. PhD thesis, Leopold Franzens Universität Innsbruck. 53 pp.

Pezzatti, G., Reinhard, M., and Conedera, M., 2010. Swissfire: die neue schweizerische Waldbranddatenbank. *Schweizerische Zeitschrift fur Forstwesen*, 161:465–469.

Pezzatti, G., Angelis, A., and Conedera, M., 2016. Potenzielle Entwicklung der Waldbrandgefahr im Klimawandel. In *Wald im Klimawandel*, pages 223–245. Haupt Verlag.

Plavcan, D. and Mayr, G. J., 2015. Towards an Alpine Foehn Climatology. In *International Conference on Alpine Meteorology*, Innsbruck, Austria.

Plavcan, D., Mayr, G. J., and Zeileis, A., 2014. Automatic and probabilistic foehn diagnosis with a statistical mixture model. *Journal of Applied Meteorology and Climatology*, 53:652–659.

Reinhard, M., Rebetez, M., and Schlaepfer, R., 2005. Recent climate change: rethinking drought in the context of forest fire research in Ticino, South of Switzerland. *Theoretical and Applied Climatology*, 82:17–25.

Richner, H. and Hächler, P., 2013. Understanding and Forecasting Alpine Foehn. In *Mountain weather research and forecasting: recent progress and current challenges*, pages 219–260. Springer.

Richner, H., Dürr, B., Gutermann, T., and 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.

Robinne, F.-N., Burns, J., Kant, P., Flannigan, M., Kleine, M., de Groot, B., and Wotton, D., 2018. Global fire challenges in a warming world. *IUFRO Occasional Paper*, 32:60 pp.

Röthlisberger, M., Sprenger, M., Flaounas, E., Beyerle, U., and Wernli, H., 2020. The substructure of extremely hot summers in the Northern Hemisphere. *Weather and Climate Dynamics*, 1:45–62.

Schroeder, M. J., Glovinsky, M., Henricks, V. F., Hood, F. C., and Hull, M. K., 1964. Synoptic weather types associated with critical fire weather. Technical report, USDA Forest Service: Pacific Southwest Forest And Range Experiment Station Berkeley CA. 503 pp.

Schuetz, J. and Steinhauser, F., 1955. Neue Föhnuntersuchungen aus dem Sonnblickgebiet. *Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B*, 6:207–224.

Sharples, J. J., 2018. Foehn winds. In *Encyclopedia of Wildfires and Wildland-Urban Interface (WUI) Fires*, pages 1–7. Springer International Publishing.

Sharples, J. J., Mills, G. A., McRae, R. H., and Weber, R. O., 2010. Foehn-like winds and elevated fire danger conditions in southeastern Australia. *Journal of Applied Meteorology and Climatology*, 49:1067–1095.

Simpson, C., Pearce, H., Sturman, A., and Zawar-Reza, P., 2014. Behaviour of fire weather indices in the 2009–10 New Zealand wildland fire season. *International Journal of Wildland Fire*, 23:1147–1164.

Sprenger, M., Dürr, B., and Richner, H., 2016. Foehn studies in Switzerland. In *From weather observations to atmospheric and climate sciences in Switzerland: Celebrating 100 years of the Swiss Society for Meteorology*, pages 215–247. vdf Hochschulverlag AG.

Sprenger, M., Schemm, S., Oechslin, R., and Jenkner, J., 2017. Nowcasting foehn wind events using the AdaBoost machine learning algorithm. *Weather and Forecasting*, 32:1079–1099.

Steinacker, R., 2006. Alpiner Föhn - eine neue Strophe zu einem alten Lied. *promet*, 32:3–10.

Swisstopo, B., 2016. Formeln und Konstanten für die Berechnung der Schweizerischen schiefachsigen Zylinderprojektion und der Transformation zwischen Koordinatensystemen. Bundesamt für Landestopographie. 20 pp.

Thomas, P. A., McAlpine, R. S., Hirsch, K., and Hobson, P., 2010. How a fire burns. In *Fire in the Forest*, page 26–53. Cambridge University Press.

Turco, M., Hardenberg, J., AghaKouchak, A., Llasat, M., Provenzale, A., and Trigo, R., 2017. On the key role of droughts in the dynamics of summer fires in Mediterranean Europe. *Scientific Reports*, 7:1–10.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., and Lamarque, J.-F., 2011. The representative concentration pathways: an overview. *Climatic change*, 109: 5–31.

Vergeiner, J., 2004. *South foehn studies and a new foehn classification scheme in the Wipp and Inn valley*. PhD thesis, University of Innsbruck. 105 pp.

Westerling, A., R.Cayan, D., Brown, T., and Riddle, L., 2004. Climate, Santa Ana Winds and Autumn Wildfires in Southern California. *Eos Transactions AGU*, 85:289–296.

Widmer, R., 1966. Statistische Untersuchungen über den Föhn im Reusstal und Versuch einer objektiven Föhnprognose für die Station Altdorf. *Vierteljahrsschrift Naturforsch. Ges. Zürich*, 111:331–375.

Wilhelm, M., Buzzi, M., Sprenger, M., and Hächler, P., 2012. COSMO-2 model performance in forecasting foehn: A systematic process-oriented verification. Master's thesis, ETH Zürich. 55 pp.

WMO, E., 1992. *International Meteorological Vocabulary*, volume 182. World Meteorological Organization. 784 pp.

Zumbrunnen, T., Bugmann, H., Conedera, M., and Bürgi, M., 2009. Linking forest fire regimes and climate - a historical analysis in a dry inner alpine valley. *Ecosystems*, 12:73–86.

Zweifel, L., Mayr, G., and Stauffer, R., 2016. Probabilistic Foehn forecasting for the Gotthard Region based on model output statistics. Master's thesis, University of Innsbruck. 81 pp.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Evaluating foehn development in a changing climate using machine learning and investigating the impact of foehn on forest fire occurrence and severity

Authored by (in block letters):

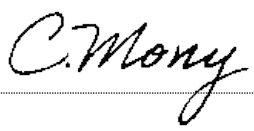
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Mony	Christoph

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
23.12.2020	

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.