



# AMERICAN METEOROLOGICAL SOCIETY

*Bulletin of the American Meteorological Society*

## **EARLY ONLINE RELEASE**

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/BAMS-D-17-0200.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Mayr, G., D. Plavcan, L. Armi, A. Elvidge, B. Grisogono, K. Horvath, P. Jackson, A. Neururer, P. Seibert, J. Steenburgh, I. Stiperski, A. Sturman, Ž. Vecenaj, J. Vergeiner, S. Vosper, and G. Zängl, 2018: The Community Foehn Classification Experiment. Bull. Amer. Meteor. Soc. doi:10.1175/BAMS-D-17-0200.1, in press.



# The Community Foehn Classification Experiment

Georg J. Mayr\*

*University of Innsbruck, Innsbruck, Austria*

David Plavcan†

*University of Innsbruck, Innsbruck, Austria*

Laurence Armi

*Scripps Institution of Oceanography, University of California San Diego, La Jolla, California*

Andrew Elvidge

*University of East Anglia, Norwich, UK*

Branko Grisogono

*University Zagreb, Zagreb, Croatia*

Kristian Horvath

*Meteorological and Hydrological Service, Zagreb, Croatia*

Peter Jackson

*University of Northern British Columbia, Prince George, Canada*

Alfred Neururer

*Central Institution for Meteorology and Geodynamics, Innsbruck, Austria*

18 Petra Seibert

19 *University of Natural Resources and Life Sciences, Vienna, Austria*

20 James W Steenburgh

21 *University of Utah, Salt Lake City, USA*

22 Ivana Stiperski

23 *University of Innsbruck, Innsbruck, Austria*

24 Andrew Sturman

25 *University of Canterbury, Christchurch, New Zealand*

26 Željko Večenaj

27 *University of Zagreb, Zagreb, Croatia*

28 Johannes Vergeiner

29 *Central Institution for Meteorology and Geodynamics, Innsbruck, Austria*

30 Simon Vosper

31 *Met Office, Exeter, UK*

32 Günther Zängl

33 *Deutscher Wetterdienst, Offenbach, Germany*

34 *\*Corresponding author address: Georg J. Mayr, Department of Atmospheric and Cryospheric Sci-*  
35 *ences, University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria.*

36 E-mail: georg.mayr@uibk.ac.at

<sup>37</sup> †Current affiliation: UBIMET, Vienna, Austria

## ABSTRACT

38 Strong winds crossing elevated terrain and descending to its lee occur over  
39 mountainous areas worldwide. Winds fulfilling these two criteria are called  
40 “foehn” in this paper although different names exist depending on region,  
41 sign of temperature change at onset, and depth of overflowing layer. They  
42 affect local weather and climate and impact society. Classification is difficult  
43 because other wind systems might be superimposed on them or share some  
44 characteristics. Additionally, no unanimously agreed-upon name, definition  
45 nor indications for such winds exist. The most trusted classifications have  
46 been performed by human experts. A classification experiment for different  
47 foehn locations in the Alps and different classifier groups addressed hitherto  
48 unanswered questions about the uncertainty of these classifications, their re-  
49 producibility and dependence on the level of expertise. One group consisted  
50 of mountain meteorology experts, the other two of Masters degree students  
51 who had taken mountain meteorology courses, and a further two of objective  
52 algorithms. Sixty periods of 48 hours were classified for foehn/no foehn at  
53 five Alpine foehn locations. The intra-human-classifier detection varies by  
54 about 10 percentage points (interquartile range). Experts and students are  
55 nearly indistinguishable. The algorithms are in the range of human classifica-  
56 tions. One difficult case appeared twice in order to examine reproducibility of  
57 classified foehn duration, which turned out to be 50% or less. The classifica-  
58 tion dataset can now serve as a testbed for automatic classification algorithms,  
59 which - if successful - eliminate the drawbacks of manual classifications: lack  
60 of scalability and reproducibility.

## 61 **1. Introduction**

62 Many processes and phenomena in the atmosphere need to be diagnosed – from low pressure  
63 systems with fronts in midlatitudes and hurricanes in the tropics, to fog or lightning. Some diag-  
64 noses are easy to make. Hearing thunder identifies lightning, and not being able to see a building  
65 less than 1 km away during daytime indicates fog. These diagnoses can even be automated with  
66 suitable instrumentation – to identify lightning from its signature in the electromagnetic waves it  
67 emits and fog from scattering of a light source. Some processes and phenomena, however, are  
68 much harder to classify, often because not enough information is available or the process itself  
69 is insufficiently understood. Lately, methods from statistics and machine learning in combination  
70 with a huge increase in computing power have been harnessed with ever increasing success to  
71 tackle more and more difficult classification tasks, earning them the label “artificial intelligence”.  
72 Arguably the largest progress has been made in classifying images, from spotting a dog on a photo  
73 to identifying a particular person. The underlying neural-network algorithms, however, typically  
74 need thousands or even hundreds of thousands of pre-classified images provided by humans in or-  
75 der to “learn”. Such “supervised” learning is much easier than “unsupervised learning” for which  
76 no “truth” exists. This is the area where classifications by human experts are still the gold stan-  
77 dard, albeit with several drawbacks: lack of scalability and reproducibility, and unknown error  
78 rates. Because only few people have the required expertise to perform a classification, which takes  
79 a substantial amount of time, the classification task cannot be extended to an arbitrarily large num-  
80 ber of instances, and comparison of classifications among different experts or by the same expert  
81 performed at different times are at best extremely rare.

82 A group of experts collaborated recently on such a task to remedy two of the classification  
83 drawbacks by providing estimates of classification uncertainty and reproducibility, and a database

84 against which existing and future algorithms can be tested. The classification task identified peri-  
85 ods of downslope windstorms in time series of weather station measurements.

86 Such windstorms result from winds that cross topographic obstacles and accelerate as they de-  
87 scend to their lee. They occur over mountainous locations worldwide and are known by differ-  
88 ent names, which are sometimes also used to refer to an additional characteristic. Because no  
89 all-encompassing name exists this article will use “foehn” for simplicity *without* implying a tem-  
90 perature increase during its onset, or a specific region. Foehn affects local weather and climate  
91 and impacts agriculture (growing conditions due to temperature and humidity changes; top soil  
92 erosion), tourism (reliable spots for wind and kite surfing), artificial snow making (change of wet-  
93 bulb temperatures), air pollution (trapping pollutants in cold pools underneath the foehn layer, or  
94 sweeping them away in case of break-through), human health (reduction of air pollution), forest  
95 fires (intensifying them to uncontrollable extents), ground traffic (toppling trucks; snow or sand  
96 drifts; blasting of vehicles with sand and small rocks), and air traffic (closure of runways when  
97 crosswinds are too high). The increasing density of automatic weather stations allows the obser-  
98 vation of such winds at progressively more locations. Classification, however, is difficult because  
99 other wind systems such as radiatively-driven downslope/downvalley winds might be superim-  
100 posed on foehn, share some of its characteristics, or because not enough information is available.  
101 The difficulty is compounded because no unanimously agreed-upon definition of foehn and its  
102 indications exist, foehn occurs in a variety of synoptic-scale and mesoscale settings, and different  
103 names are being used depending on region, the sign of temperature change at its onset, and its  
104 depth.



## 105 **2. Classification task**

106 Nevertheless, two unanimously agreed-upon characteristics are that air crosses an obstacle and  
107 that it descends and accelerates on the downwind side causing strong winds. A fairly simple  
108 conceptual model of the flow situation after the onset of foehn, corroborated by field campaigns,  
109 laboratory experiments, computer simulations, and theoretical investigations, is shown in Fig.  
110 1. Unfortunately, no continuous measurements covering the vertical cross-section are routinely  
111 available for classification; only weather stations at the ground. Nowadays with the proliferation  
112 of automatic weather stations and mesonets in some regions, measurement(s) close to the crest of  
113 the obstacle are also available so that the first foehn characteristic of air *crossing* the topographic  
114 obstacle can be checked. The second characteristic that air *descends* leads to adiabatic warming  
115 and consequentially to a decrease in relative humidity. It can be examined through differences  
116 between crest and downwind station of variables which are approximately conserved in foehn  
117 flow, such as potential temperature or mixing ratio.

118 Classification is made more ambiguous by processes for which potential temperature and mixing  
119 ratio are not conserved, i.e. turbulent mixing within the foehn flow, at the surface and its upper  
120 interface; mixing air in from tributaries; phase changes of water (formation and evaporation of  
121 liquid and solid particles); daytime warming and nighttime cooling due to surface sensible heat  
122 flux. How large these diabatic effects are varies with season, time of day, location, and large-scale  
123 and mesoscale flow configurations. Information about their contribution is not readily available so  
124 that classifications become difficult and possess an unknown and variable degree of uncertainty.

## 125 **3. The community foehn classification experiment**

126 The community foehn classification experiment set out to quantify the uncertainty of human  
127 foehn classifications, to compare them to machine classifications and to provide a data set for the

128 development of foehn classification algorithms. Three groups of human experts and two objective  
129 algorithms faced the task of identifying foehn periods. The first group (most of them are co-authors  
130 of this paper) consisted of 26 seasoned experts in mountain meteorology from different continents  
131 with operational or research backgrounds and thus a broad range of concepts of what constitutes  
132 foehn. The other two groups are students taking the advanced weather forecasting course at the  
133 University of Innsbruck in 2016 (34) and 2017 (18), respectively. The student groups had a fairly  
134 homogeneous level of expertise because they had received four hours of lectures on foehn and had  
135 to apply it in homework problems in their advanced weather forecasting course. It was explained  
136 to the students why it was crucial for the outcome of this study that they worked completely inde-  
137 pendently. In addition to human experts, two algorithms were used that also employ the concept  
138 shown in Fig. 1. One, labeled A1 henceforth, is in operational use by the Swiss weather service.  
139 It uses percentiles of the distribution of the difference of potential temperature between crest and  
140 downstream locations (small, cf. Fig. 1), wind speed (high) and relative humidity (low) as hard  
141 thresholds for the classification of three categories: no foehn, foehn air mixed with cold valley air,  
142 and foehn. The second algorithm, A2, in operational use at the University of Innsbruck, learns  
143 from the data by itself and does not use hard thresholds. It uses so-called statistical mixture mod-  
144 els to fit two or more parametric distributions to the observed distribution of classifying variables,  
145 such as potential temperature difference between crest and downwind stations, and wind speed,  
146 to yield a probability for foehn between 0 and 1, instead of merely a binary yes/no classification.  
147 Both algorithms require that the appropriate directional sector for foehn winds be manually set.

148 The classification experiment was designed to strike a balance between ideal goals and practical  
149 feasibility for the human classifiers. Therefore, five topographically different locations of differ-  
150 ing annual foehn frequency in the Swiss Alps were selected (Table 1 and Fig. 2). Twelve 48-hour  
151 periods at each station yielded a total of 60 cases, for which the experts had to classify south foehn

152 periods lasting at least 1 h at 30-minute resolution. One of the co-authors, who did not himself  
153 manually classify (D. Plavcan), selected these cases based on results from the two automated clas-  
154 sification algorithms, A1 and A2, to cover all permutations: phases of foehn/no-foehn for which  
155 both, only one, or none agreed. Cases contained none, one or several foehn periods, respectively.  
156 Unbeknownst to the classifiers, one difficult 48-hour period appeared twice in order to estimate  
157 reproducibility.

158 Each participant received a wind-speed-coded wind rose for each location, a pseudo-3D image  
159 of the location from Google Earth, exact coordinates, plots of meteorological variables for each  
160 of the 60 periods of 48 hours, and instructions that contained an annotated example of an addi-  
161 tional case reproduced here in Fig. 3. To classify only south foehn events, air had to cross the  
162 Alpine crest from south to north as indicated by wind direction at the crest plotted in black instead  
163 of gray, which is fulfilled for the whole 48-hour-period in this case. Three periods of foehn are  
164 inferred; from 9:00—10:20, 11:10—14:30 and 31:00—45:20 (as hh:min). During these periods,  
165 similar potential temperatures at crest and the classification location imply the second foehn char-  
166 acteristic of lee-slope descent. Wind directions are from the appropriate sector<sup>1</sup> and wind speeds  
167 are higher. Temperatures increase at the onset of each period, presumably when foehn erodes an  
168 underlying shallow cold pool. Humidity also drops, reflecting the draw-down of drier air from  
169 higher altitudes. Because *relative* humidity (%) instead of specific humidity (g/kg) is plotted, the  
170 temperature increase additionally contributes to a drop in relative humidity.

#### 171 **4. Results**

172 The three human groups classified foehn duration during the 12 x 48-h periods at each of the  
173 five locations broadly similarly as Fig. 4 shows. Median durations (colored horizontal lines) are

---

<sup>1</sup> deduced from wind roses and topography maps; not shown

174 within a few percentage points of each other. The group of mountain meteorology experts have the  
175 most diverse backgrounds and consequently concepts of what constitutes foehn. As a result, their  
176 classification variation is larger than that of the second group of students who all had the same  
177 foehn concept instilled in their course. The variation of the first group of students, on the other  
178 hand, is larger; mainly because of a few outliers at each location.

179 The variation and thus classification uncertainty is smallest at location 4, a station at the northern  
180 edge of the Alps. The largest uncertainty occurred for location 1, where foehn can potentially blow  
181 from several wind sectors and for which the crest station might not always be representative of the  
182 upstream conditions.

183 The agreement between the algorithms and human classifications varies. A1 is within a few  
184 percentage points of the medians of the human groups at locations 2 and 3; A2 at locations 1 and  
185 4. However, they are at the margins of human classifications for locations 2 (A2), 3 (A2) and 5  
186 (A1 and A2), and A1 is even outside at locations 1 and 4.

#### 187 *a. Classification example*

188 Fig. 5 shows the classifications from the three groups of human classifiers and the two algorithms  
189 for one of the 60 cases. At about midday of the *second* day potential temperature at the valley  
190 station 1 reached a value close to that of the crest station (purple line), indicating descent of  
191 air. Wind speeds also increased. In the evening the signals in the variables reverse, indicating  
192 the cessation of foehn conditions. Human classifications agree on a core period of foehn from  
193 11:00-14:30 (labeled “easy” in the figure) but differ in onset and end times, with end times less  
194 unanimous than onset times. The two algorithms classify similarly.

195 The nighttime period between day’s 1 and 2, on the other hand, is more difficult. About 60%  
196 of the experts and students classified it as foehn (labeled “difficult”), again agreeing for the core

197 period but differing for onset and even more so for end times. On the evening of the first day the  
198 wind direction changed into the foehn sector. At the same time both average and peak wind speeds  
199 increased and potential temperature also increased. Unlike the “easy” period, however, potential  
200 temperature is five Kelvin colder than at the crest, which likely led the other 40% to classify it as  
201 a radiatively cooled nocturnal downslope/downvalley flow. Air originating from a different level  
202 than represented by the crest station (cf. Fig. 2) and mixing of foehn air with radiatively cooled  
203 air from the valley and its tributaries might have been responsible for such a large difference. The  
204 three-category algorithm A1 classifies no foehn, whereas the mixture model algorithm A2 gives a  
205 probability close to 1 that it is foehn. The decrease and fluctuations of the probability towards the  
206 end of the period stems from the decrease and fluctuations in wind speed and later on the increase  
207 in potential temperature difference.

208 This “difficult” period indicates that a simple “yes” or “no” might not be enough for all ap-  
209 plications when it comes to classifying foehn flows, for example because of the superposition of  
210 foehn and a radiatively cooled downvalley wind. Algorithm A1 adds the third category of “mixed  
211 foehn/valley air” (although it does not classify it as such in this particular case). Algorithm A2  
212 gives a continuous probability of foehn occurrence.

### 213 *b. Changes in classification uncertainty*

214 Over all 60 cases, delineating the beginning and end of a foehn event had a higher variability  
215 among all classifiers. Although the majority of classified foehn events started with a tempera-  
216 ture increase, uncertainty was not clearly different from events which started with no change or  
217 a decrease of temperature. Classification uncertainty was also higher for nighttime than daytime  
218 for similar reasons as in the “difficult” period of Fig. 5. Classification uncertainty also varied  
219 somewhat seasonally with low uncertainty in fall (SON) and winter (DJF) months, highest uncer-

220 tainty in spring (MAM) particularly among human classifiers, and medium uncertainty in summer  
221 months (JJA).

### 222 *c. Reproducibility*

223 To evaluate reproducibility, one of the more difficult cases (at location 1) occurred twice in  
224 the data set, unbeknownst to the classifiers. Fig. 6 shows the relative frequency of the absolute  
225 difference of foehn duration classified at the first occurrence and the second occurrence of that  
226 case. Ideally and for perfect reproducibility, the difference in classified foehn duration among the  
227 identical cases is zero. However, fewer than half of the classifiers achieved perfect reproducibility.

228 This lack of reproducibility is worrisome although probably less extreme for easier cases. Nev-  
229 ertheless, it corroborates the first author's personal experience of classifying foehn at different  
230 locations globally.

### 231 *d. Dataset*

232 The dataset will be available at UC Irvine, which hosts a large repository of classification data  
233 sets, at <https://archive.ics.uci.edu/ml/about.html>.

## 234 **5. Conclusion**

235 Several lessons have been learned from this experiment that add on the one hand supporting  
236 evidence to what was previously at least informally known from other classification tasks ((i)-(iii),  
237 and on the other hand ((iv) - (vi)) add new knowledge: (i) Busy experts are willing to volunteer  
238 a chunk of their scarce time provided the classification task is an intellectually challenging puz-  
239 zle; (ii) Human experts use implicit (and in the case of the Masters students explicitly taught)  
240 physically-based concepts to help them distinguish between the two categories of foehn/no foehn;

241 (iii) Expert classifications carry uncertainty and are not even necessarily reproducible, which needs  
242 to be quantified (as here) or at least considered when interpreting results using such classifica-  
243 tions; (iv) Uncertainty is largest for onset and even more so for the ending of a foehn event and  
244 also larger during the night; (v) Combining advanced statistical and/or machine learning models  
245 with physically-based concepts for choosing their input variables yields similar results to those of  
246 human experts. In addition, they easily scale to longer time series or more locations and are re-  
247 producible, which is a fundamental scientific requirement and allows the comparison of different  
248 data sets (foehn occurrence at different locations in this case). It is thus highly recommended to  
249 develop objective classification procedures, ideally without having to resort to manually specified  
250 and/or hard limits. If the algorithms are additionally made available as packages of open-source  
251 languages, foehn classifications can easily be reproduced by other researchers; (vi) Diagnoses  
252 contain more information when they are probabilistic instead of binary yes/no – a concept that has  
253 already been implemented for a long time in (weather) forecasts.

254 In addition to shedding light on human and machine classification of foehn, the dataset allows  
255 the testing of existing and newly developed algorithms for unsupervised learning tasks when truth  
256 is not known, such as in the case of foehn occurrence. It can also serve a community interested in  
257 estimating the accuracy of previous human foehn classifications and climatologies.

258 *Acknowledgments.* Many thanks to Achim Zeileis for discussions how to best design the exper-  
259 iment. A huge thanks to the experts and the 2016 and 2017 cohorts of the Advanced Weather  
260 Forecasting class who classified these 60 periods.

## 261 **References**

262 Brinkmann, W. A. R., 1971: What is a foehn? *Weather*, **26 (6)**, 230–240, doi:10.1002/j.1477-8696.  
263 1971.tb04200.x.

264 Dürr, B., 2008: Automatisiertes Verfahren zur Bestimmung von Föhn in Alpentälern. Tech.  
265 Rep. 223, Arbeitsberichte der MeteoSchweiz, 22 pp. URL [http://www.meteoschweiz.admin.ch/  
266 content/dam/meteoswiss/de/Ungebundene-Seiten/Publikationen/Fachberichte/doc/ab223.pdf](http://www.meteoschweiz.admin.ch/content/dam/meteoswiss/de/Ungebundene-Seiten/Publikationen/Fachberichte/doc/ab223.pdf).

267 Elvidge, A. D., and I. A. Renfrew, 2016: The causes of foehn warming in the lee of mountains. *Bul-  
268 letin of the American Meteorological Society*, **97** (3), 455–466, doi:10.1175/bams-d-14-00194.  
269 1.

270 Jackson, P. L., G. Mayr, and S. Vosper, 2013: Dynamically-driven winds. *Mountain  
271 Weather Research and Forecasting: Recent Progress and Current Challenges*, F. K. Chow,  
272 S. F. J. de Wekker, and B. J. Snyder, Eds., Springer, chap. 3, 121–218, doi:10.1007/  
273 978-94-007-4098-3\_3.

274 Leisch, F., 2004: FlexMix: A general framework for finite mixture models and latent class regres-  
275 sion in R. *Journal of Statistical Software*, **11** (8), 1–18, doi:10.18637/jss.v011.i08.

276 Plavcan, D., G. J. Mayr, and A. Zeileis, 2014: Automatic and probabilistic foehn diagno-  
277 sis with a statistical mixture model. *J. Appl. Meteor. Climatol.*, **53**, 652–659, doi:10.1175/  
278 jamc-d-13-0267.1.

279 Richner, H., and P. Hächler, 2013: Understanding and forecasting Alpine foehn. *Mountain  
280 Weather Research and Forecasting: Recent Progress and Current Challenges*, F. K. Chow,  
281 S. F. J. de Wekker, and B. J. Snyder, Eds., Springer, 219–260, doi:10.1007/978-94-007-4098-3\_  
282 4.

283 Seibert, P., 2012: The riddles of foehn - introduction to the historic articles by Hann and Ficker.  
284 *Meteorologische Zeitschrift*, **21** (6), 607–614, doi:10.1127/0941-2948/2012/0398.



285 von Ficker, H., 1910: On the formation of foehn winds on the northern side of the Alps (translated  
286 and edited by Volken, E. and S. Brönnimann). *Meteorologische Zeitschrift*, **21** (6), 597–605,  
287 doi:10.1127/0941-2948/2012/0532.

288 **LIST OF TABLES**

289 **Table 1.** Weather station locations used for foehn classification with their long-term  
290 foehn frequencies determined from automatic algorithms A1 and A2. . . . . 18

291 TABLE 1. Weather station locations used for foehn classification with their long-term foehn frequencies  
 292 determined from automatic algorithms A1 and A2.

location	latitude N	longitude E	altitude [m MSL]	frequency from A1[%]	from A2 [%]	town
1	46.30287	7.84294	639	6	10	Visp
2	46.88702	8.62181	438	5	5	Altdorf
3	47.12745	9.51753	457	4	4	Vaduz
4	47.42546	9.39847	776	2	2	St. Gallen
5	47.03643	8.30097	457	<1	<1	Luzern
c (crest)	46.65346	8.61625	2287	-	-	Guetsch

293 **LIST OF FIGURES**

294 **Fig. 1.** Conceptual model of well-established foehn flow (dark grey shading) along a vertical cross-  
 295 section exhibiting the two core characteristics of air crossing the obstacle (black), which  
 296 can be ridge-like, a strait, or a pass, and descending to its lee. Flow approximately along  
 297 isentropes (thin lines); straight arrows indicate wind speed, curved ones turbulence. Colored  
 298 dots are exemplary weather station locations at crest and downwind (c.f. Table 1). . . . . 20

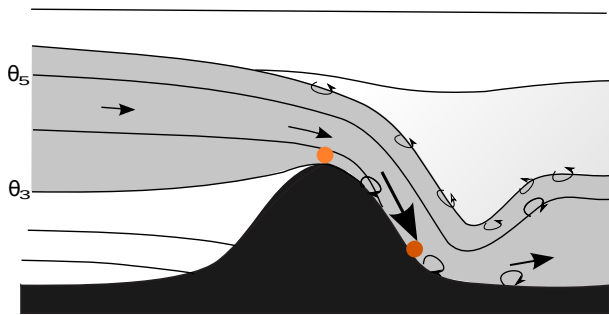
299 **Fig. 2.** Topography (m msl) and location of stations for which foehn was classified (cf.  
 300 Table 1). Measurements at the crest location (C) were used to assist clas-  
 301 sification at all locations. Digital elevation model at 250 m horizontal res-  
 302 olution from SRTM ([http://gisweb.ciat.cgiar.org/TRMM/SRTM\\_Resampled\\_](http://gisweb.ciat.cgiar.org/TRMM/SRTM_Resampled_250m/)  
 303 [250m/](http://gisweb.ciat.cgiar.org/TRMM/SRTM_Resampled_250m/)[gisweb.ciat.cgiar.org/TRMM/SRTM\\_Resampled\\_250m/](http://gisweb.ciat.cgiar.org/TRMM/SRTM_Resampled_250m/)). . . . . 21

304 **Fig. 3.** Annotated time series of an additional case at location 2 (Altdorf) supplied to classifiers with  
 305 the instruction package and other material. Upper panel: wind speed (magenta) and direction  
 306 (degrees from N; in grey but bold black when from foehn sector) at the crest station (2287  
 307 m amsl); center panel: potential temperature at crest station (magenta) and classification  
 308 location (blue) and relative humidity at classification location (green shaded); lower panel:  
 309 wind speed (blue), gusts (light blue) and direction (black; degrees from N) at location 2.  
 310 All values (except gusts) are averages over the previous 10 minutes. A hypothetical but not  
 311 unreasonable classification of three foehn episodes at the station Altdorf is marked by orange  
 312 rectangles (9:00-10:20; 11:10-14:30; 31:00-45:20). Foehn episodes had to be classified at  
 313 a resolution of complete half hour segments and a minimum duration of 1 hour. In this  
 314 example, foehn was classified between 9.0 – 10.0, 11.5 – 14.5, and 31.0 – 45.0. . . . . 22

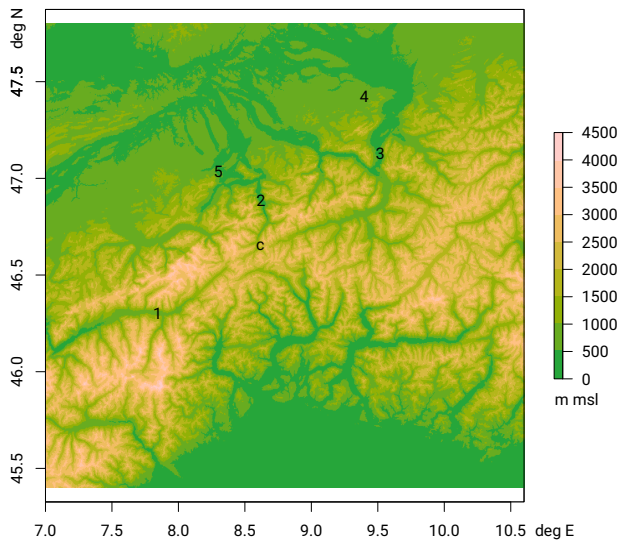
315 **Fig. 4.** Beanplots of classified foehn duration at each location relative to the total duration of the  
 316 time series of 12 times 48 hours stratified by classifier groups: experts, two Masters student  
 317 groups, and the two algorithms. (A1 for foehn mixed with valley air and pure foehn com-  
 318 bined; for A2 a threshold of foehn probability of at least 50% is used). Black lines indicate  
 319 individual classifications, colored lines the median of each group. Areas are the empirical  
 320 densities of each group. . . . . 23

321 **Fig. 5.** Classification case at location 1. First panel: potential temperature (blue) and relative humid-  
 322 ity (green) with added potential temperature at crest (purple). Second panel: wind direction  
 323 (black), wind speed average (dark blue) and gusts (light blue) at location 1. Third panel:  
 324 proportion of human classifier groups that classified foehn during the time series. Fourth  
 325 panel: classifications with the three-category algorithm A1 (no foehn (0), foehn mixed with  
 326 valley air (1), foehn (2)); Last panel: probability of foehn from the statistical mixture model  
 327 A2. . . . . 24

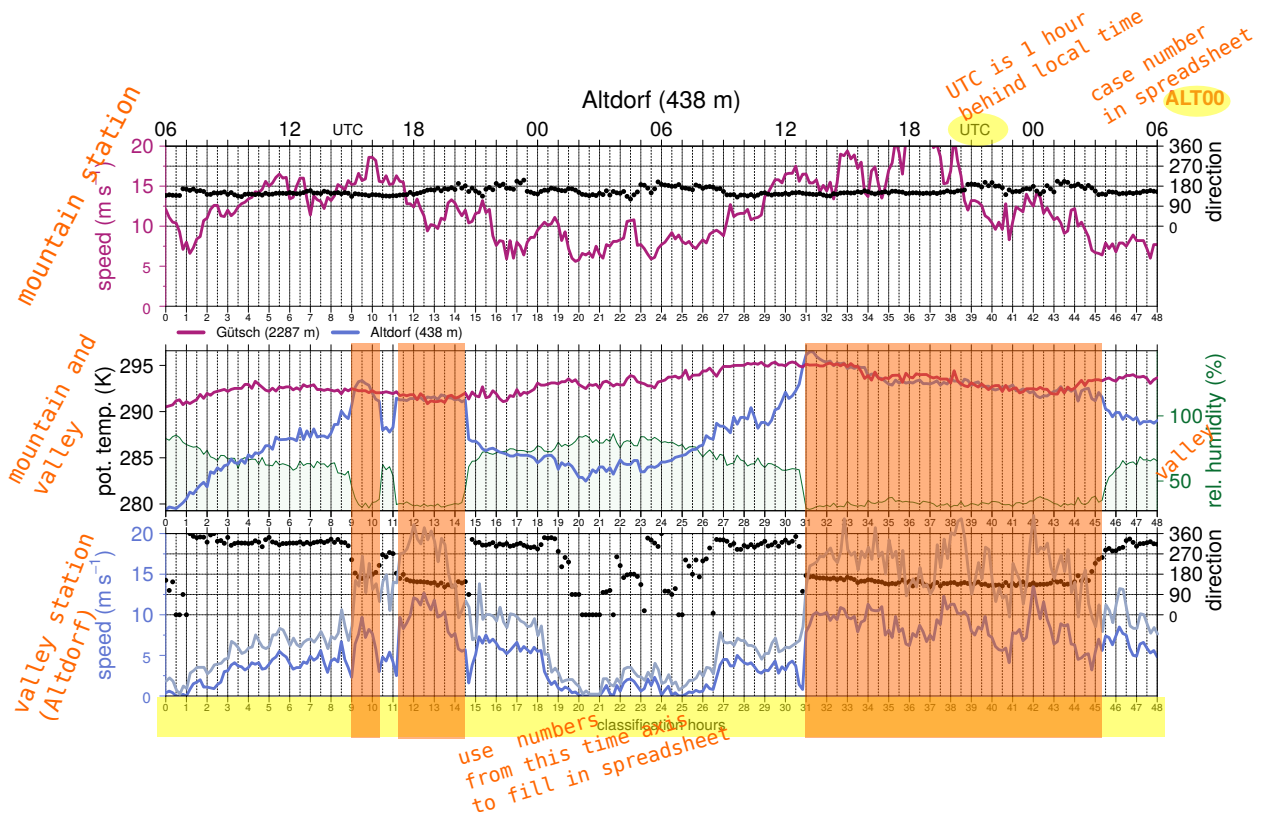
328 **Fig. 6.** Histogram of absolute difference in classified foehn duration (hours) between two identical  
 329 cases. For perfect reproducibility all classifiers should have had 0 h difference. The bars are  
 330 for the hour prior to and including the labeled duration difference. . . . . 25



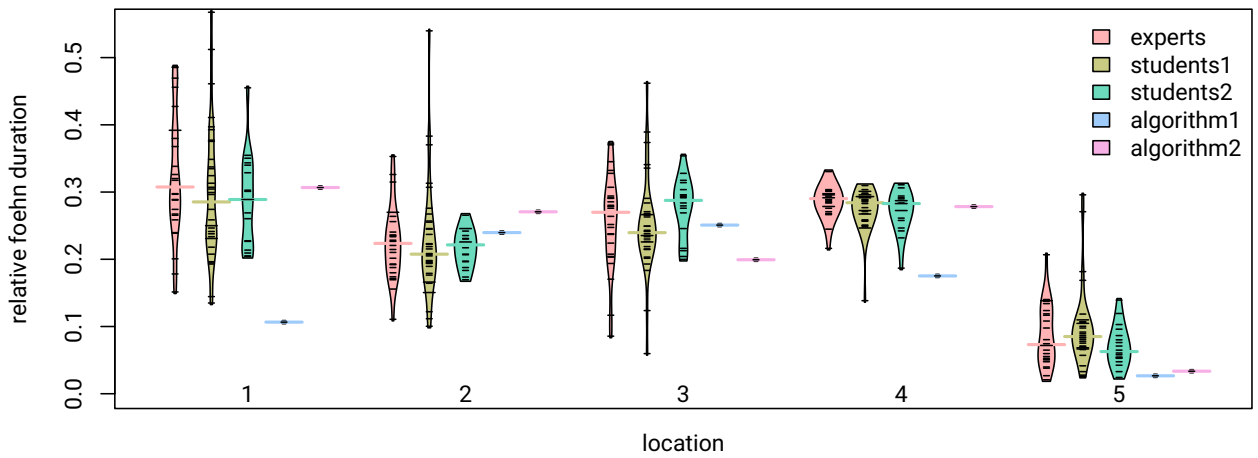
331 FIG. 1. Conceptual model of well-established foehn flow (dark grey shading) along a vertical cross-section  
 332 exhibiting the two core characteristics of air crossing the obstacle (black), which can be ridge-like, a strait, or a  
 333 pass, and descending to its lee. Flow approximately along isentropes (thin lines); straight arrows indicate wind  
 334 speed, curved ones turbulence. Colored dots are exemplary weather station locations at crest and downwind (c.f.  
 335 Table 1).



336 FIG. 2. Topography (m msl) and location of stations for which foehn was classified (cf. Table 1). Mea-  
 337 surements at the crest location (C) were used to assist classification at all locations. Digital elevation model  
 338 at 250 m horizontal resolution from SRTM ([http://gisweb.ciat.cgiar.org/TRMM/SRTM\\_Resampled\\_](http://gisweb.ciat.cgiar.org/TRMM/SRTM_Resampled_250m/)  
 339 [250m/gisweb.ciat.cgiar.org/TRMM/SRTM\\_Resampled\\_250m/](http://gisweb.ciat.cgiar.org/TRMM/SRTM_Resampled_250m/)).

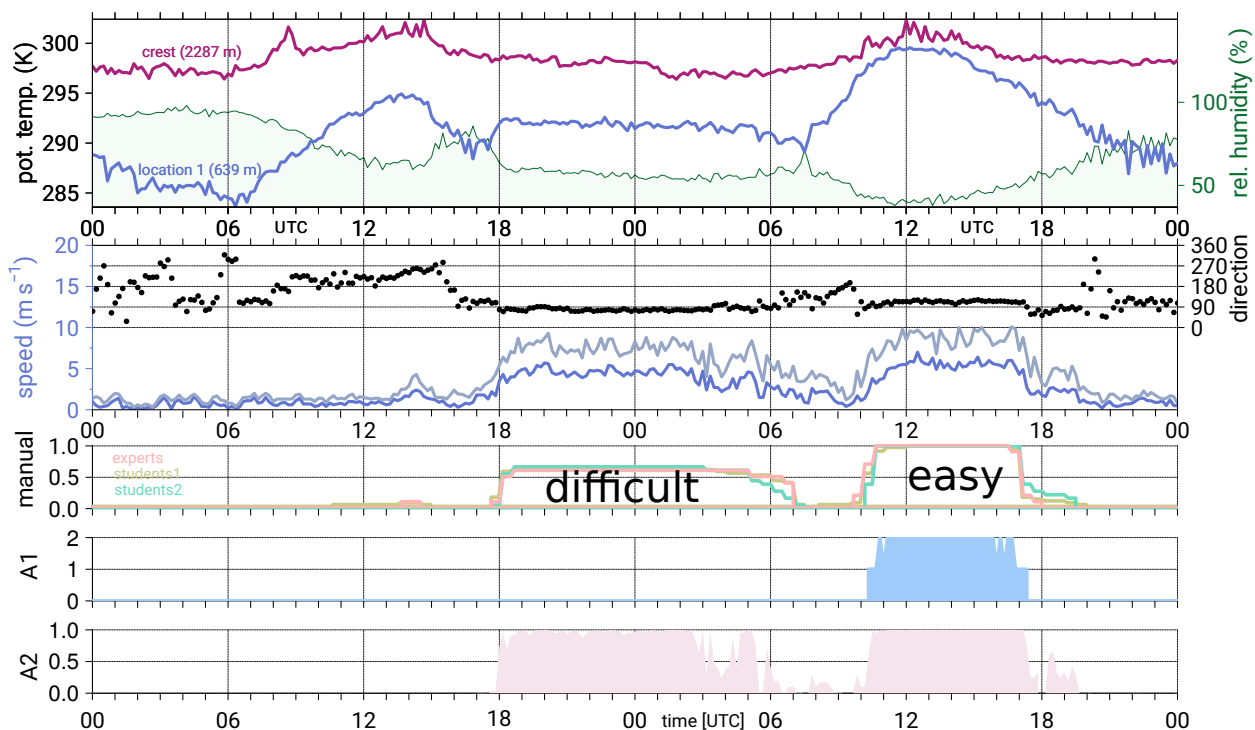


340 FIG. 3. Annotated time series of an additional case at location 2 (Altdorf) supplied to classifiers with the  
 341 instruction package and other material. Upper panel: wind speed (magenta) and direction (degrees from N;  
 342 in grey but bold black when from foehn sector) at the crest station (2287 m amsl); center panel: potential  
 343 temperature at crest station (magenta) and classification location (blue) and relative humidity at classification  
 344 location (green shaded); lower panel: wind speed (blue), gusts (light blue) and direction (black; degrees from  
 345 N) at location 2. All values (except gusts) are averages over the previous 10 minutes. A hypothetical but not  
 346 unreasonable classification of three foehn episodes at the station Altdorf is marked by orange rectangles (9:00-  
 347 10:20; 11:10-14:30; 31:00-45:20). Foehn episodes had to be classified at a resolution of complete half hour  
 348 segments and a minimum duration of 1 hour. In this example, foehn was classified between 9.0 – 10.0,  
 349 14.5, and 31.0 – 45.0.

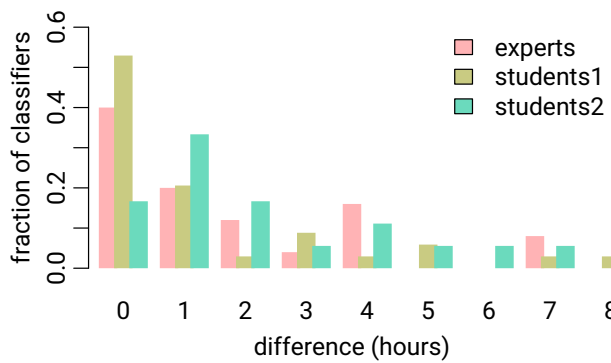


350 FIG. 4. Beanplots of classified foehn duration at each location relative to the total duration of the time series  
 351 of 12 times 48 hours stratified by classifier groups: experts, two Masters student groups, and the two algorithms.  
 352 (A1 for foehn mixed with valley air and pure foehn combined; for A2 a threshold of foehn probability of at least  
 353 50% is used). Black lines indicate individual classifications, colored lines the median of each group. Areas are  
 354 the empirical densities of each group.





355 FIG. 5. Classification case at location 1. First panel: potential temperature (blue) and relative humidity (green)  
 356 with added potential temperature at crest (purple). Second panel: wind direction (black), wind speed average  
 357 (dark blue) and gusts (light blue) at location 1. Third panel: proportion of human classifier groups that classified  
 358 foehn during the time series. Fourth panel: classifications with the three-category algorithm A1 (no foehn (0),  
 359 foehn mixed with valley air (1), foehn (2)); Last panel: probability of foehn from the statistical mixture model  
 360 A2.



361 FIG. 6. Histogram of absolute difference in classified foehn duration (hours) between two identical cases.  
 362 For perfect reproducibility all classifiers should have had 0 h difference. The bars are for the hour prior to and  
 363 including the labeled duration difference.