Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

MeteoSwiss

ETH zürich

Department of Environmental Systems Science

Institute for Environmental Decisions

Technical Report MeteoSwiss No. 282

Automatic Generation of Winter Storm Warnings

Robert S. Blass, Chahan M. Kropf, Irina Mahlstein, David N. Bresch

ISSN: 2296-0058 **doi:** 10.18751/pmch/tr/282.winterstormwarnings/1.0

Technical Report MeteoSwiss No. 282

Automatic Generation of Winter Storm Warnings

Robert Blass, Irina Mahlstein, David N. Bresch, Chahan Michael Kropf

Recommended citation:

Blass R., IMahlstein I., Bresch D.N. and Kropf C.M.: 2022, Automatic Generation of Winter Storm Warnings *Technical Report MeteoSwiss,* **282,** 42 pp. doi:10.18751/pmch/tr/282.winterstormwarnings/1.0

Editor:

Federal Office of Meteorology and Climatology, MeteoSwiss, © 2022

MeteoSwiss
Operation Center 1
CH-8044 Zürich-Flughafen
T +41 58 460 99 99
www.meteoschweiz.ch

Abstract

It is a task of MeteoSwiss to warn the public and authorities of upcoming extreme weather, in particular of windstorms. The current warning system is based on fixed geographical warning regions which are manually classified into levels of danger by weather forecasters. This technical report describes how to automatize the generation of geographically variable warnings for extreme weather events. An algorithm is designed and implemented to automatically produce and classify event-based warning polygons. The inputs to the algorithm are the maximum wind gusts computed by the COSMO2-E numerical weather prediction model. The data is processed with the image processing technique *Morphological Filtering* to reduce the heterogeneity in the wind speed data and to obtain warning polygons. Evaluated on past winter storms, the algorithm's *probability of detection* has improved by 0.11 to 0.48 while the *false alarm ratio* has increased by 0.1 to 0.85, compared to the preprocessed COSMO2-E wind speed output. Therefore, the implemented algorithm can form contiguous warning polygons while preserving the COSMO2-E model's accuracy. The algorithm can easily be tuned to generate warnings for other numerical weather predictions.

Abstract

Es ist die Aufgabe der MeteoSchweiz, die Bevölkerung und die Behörden vor drohenden Wetterextremen zu warnen. Das aktuelle Warnsystem basiert auf festen geografischen Warnregionen, die von Meteorologen manuell in Gefahrenstufen eingeteilt werden. Dieser technische Bericht beschreibt, wie die Erstellung von Warnungen vor Winterstürmen automatisiert werden kann. Es wird ein Algorithmus entworfen und implementiert, der automatisch ereignisbasierte Warnpolygone formt und klassifiziert. Der Algorithmus verwendet die vom numerischen Wettervorhersagemodell COSMO2-E berechneten maximalen Windgeschwindigkeiten. Die Daten werden mit der Bildverarbeitungstechnik *Morphologische Filterung* verarbeitet, um die Heterogenität in den Windgeschwindigkeitsdaten zu reduzieren und Warnpolygone zu generieren. Bei der Auswertung vergangener Winterstürme hat sich die Erkennungswahrscheinlichkeit (probability of detection) des Algorithmus im Vergleich zu den vorverarbeiteten COSMO2-E Windgeschwindigkeitsdaten um $0,11$ auf $0,48$ verbessert, während die Fehlalarmquote (false alarm ratio) um $0,1$ auf $0,85$ gestiegen ist. Der implementierte Algorithmus bildet zusammenhängende Warnpolygone, während die Genauigkeit des COSMO2-E Modells erhalten bleibt. Der Algorithmus kann weiterentwickelt werden, um die Heterogenität anderer numerischer Wettervorhersagen zu reduzieren und Warnpolygone zu bilden.

Contents

Contents vii

1 Introduction 1
- 1.1 Winter Storms in Switzerland . . . 1
- 1.2 Numerical Weather Prediction Model . . . 1
- 1.3 Winter Storms Warnings . . . 1
  - 1.3.1 Current Warning System . . . 2
  - 1.3.2 Automatic Generation of Variable Warnings . . . 4
- 1.4 Historical Windstorm Data . . . 4
- 1.5 Related Work . . . 5

2 Methodology 7
- 2.1 Weather Data Preprocessing . . . 7
  - 2.1.1 Overview . . . 7
  - 2.1.2 Temporal Constraint . . . 8
  - 2.1.3 Grouping Ensemble Members . . . 9
  - 2.1.4 Warning Levels per Grid Point . . . 9
- 2.2 Algorithm to Generate Warning Polygons . . . 9
  - 2.2.1 Filtering . . . 9
  - 2.2.2 Algorithm's Design . . . 11
- 2.3 Scores . . . 12
  - 2.3.1 Probability of Detection and False Alarm Ratio . . . 14
  - 2.3.2 Symmetric Extremal Dependence Index . . . 14
- 2.4 Differences in Warning Generations . . . 15

3 Results 16
- 3.1 Automatically Generated Winter Storm Warnings . . . 16
  - 3.1.1 Winter Storm on 30 Oct. 2018 . . . 16
  - 3.1.2 Winter Storm on 14 Dec. 2019 . . . 17
  - 3.1.3 Winter Storm on 04 Feb. 2020 . . . 18
- 3.2 Statistical Analysis . . . 20
  - 3.2.1 Warning level 1 . . . 20
  - 3.2.2 Warning level 2 . . . 20
  - 3.2.3 Warning level 3 . . . 21
  - 3.2.4 Statistical Analysis over all Warning Levels . . . 22

3.3 Feedback Weather Forecasters 24

4 Discussion and Outlook 25
4.1 Discussion of Hypotheses 25
4.2 Accuracy of Algorithmically Generated Warnings 25
4.3 Towards Fully Automatic Warnings 26

A Warning generation algorithm 28

B Abbreviations 30

Bibliography 31

Chapter 1

Introduction

1.1 Winter Storms in Switzerland

Winter storms are one type of windstorms which occur in Switzerland between October and April. Their physical origin lies in large differences in temperature and pressure between moving weather fronts over Europe (Planat 2021). Mostly, the storms move from West to East, and can have hurricane-like and spatially extensive effects (Planat 2021). The impact of a winter storm can affect all of Switzerland but often causes more damage in the North and West of the country (Planat 2021). Winter storms can last from one to a few days.

Winter storms are the most destructive storms occurring in Switzerland. The strong wind gusts, which can reach up to double the average wind speed (up to $250\ km/h$), are the main reason for their damaging effect (Planat 2021). Indeed, a doubling in wind speed corresponds to a quadrupling in wind force (Planat 2021). Wind gusts with large forces can cause uprooted trees, damaged buildings, disrupted electricity supplies, or inference with public transport (Planat 2021).

1.2 Numerical Weather Prediction Model

Weather forecasting is used to forecast future weather conditions. At MeteoSwiss, weather forecasting is based on *Numerical Weather Predictions* (NWP). MeteoSwiss developed its own NWP model COSMO2-E (COSMO: Consortium for Small-scale Modeling) which forecasts upcoming weather situations (MeteoSchweiz 2022). Weather variables of interest are computed at grid points of a grid with a resolution of $2.1 \times 2.1\ km$. The model comprises an ensemble of 21 forecasts, called ensemble members (MeteoSchweiz 2022). The model can forecast upcoming hazardous weather conditions, such as winter storms, up to five days ahead of the event.

1.3 Winter Storms Warnings

It is crucial to warn people and authorities of Switzerland of upcoming extreme weather conditions to reduce damages and fatalities. Warning of an upcoming danger allows taking measures to reduce its impact. A duty of MeteoSwiss is to issue warnings of upcoming winter storms. Generating and communicating warnings is a complex task, as

numerous decisions must be made. This task includes the following (non-exhaustive list) points:

- Defining levels of danger. The warnings are issued in these levels.
- Defining spatial regions or polygons. This enables the mapping of warning levels to contiguous areas of Switzerland.
- Setting start and end time of a warning. This shows in which time interval hazardous conditions are expected.
- Setting a height dependency, since a storm can affect only areas above a certain height. This is optional.

After issuing a warning, a verification of said warning needs to be undertaken. A perfect warning does contain neither misses (no observed winter storm for which no warning is issued) nor false alarms (no warning for a winter storm that does not occur). Issuing a perfect warning of winter storms for Switzerland is impossible due to the accuracy limitations in NWP models and the mountainous topology of Switzerland. Thus, in most cases, there is a trade-off between the miss rate and the false alarm rate.

1.3.1 Current Warning System

	Altitude	Level 1	Level 2	Level 3	Level 4	Level 5
Lowland/ Jura	$< 1000m$	*0km/h*	*70km/h*	*90km/h*	*110km/h*	*140km/h*
	$\geq 1000m$	*0km/h*	*100km/h*	*130km/h*	*160km/h*	*200km/h*
Pre-Alps/ Alps, Alp South Side	$< 1600m$	*0km/h*	*70km/h*	*90km/h*	*110km/h*	*140km/h*
	$\geq 1600m$	*0km/h*	*100km/h*	*130km/h*	*160km/h*	*200km/h*

Table 1.1: Warning levels in terms of maximum wind speed thresholds defined by MeteoSwiss. Different thresholds are applied for *Lowland/ Jura* and *Pre-Alps/ Alps, and Alp South Side*, as well as for low and high altitudes (in $meters\ above\ mean\ sea\ level$) (as defined in (MeteoSchweiz 2021), pg. 6, Table 3).

The current (2022) warning system for winter storms applied by MeteoSwiss considers five warning levels: level 1 (no hazard), level 2 (moderate hazard), level 3 (significant hazard), level 4 (severe hazard), and level 5 (very severe hazard), as seen in Table 1.1 (the shown coloring scheme is applied throughout this report). The warning levels are defined in terms of wind speed thresholds and are set differently for low and high altitudes. Thresholds also vary between the *Lowland/ Jura* and *Pre-Alps/ Alps, and Alp South Side*. Applying different thresholds for low and high altitudes allows to account for mountainous areas where higher wind gusts are expected more often. Because *Lowland/ Jura* is less mountainous than *Pre-Alps/ Alps, and Alp South Side*, the altitude threshold is set to $1000\ meters\ above\ mean\ sea\ level\ (mMSL)$ for the former, and $1600\ mMSL$ for the latter. The two areas are shown in in Figure 1.1, whereby *Lowland/ Jura* is colored blue and *Pre-Alps/ Alps, and Alp South Side* beige.

Further, Switzerland is spatially divided into fixed warning regions, shown in red in Figure 1.1. These regions were formed by weather experts and are based on river catchments. For winter storm warnings, MeteoSwiss weather forecasters assign a warning level to every warning region. During this process, weather forecasters use their knowledge and expertise of extreme weather situations, combined with the outputs from the NWP. An example of such a warning generated by weather forecasters is shown in Figure 1.2.

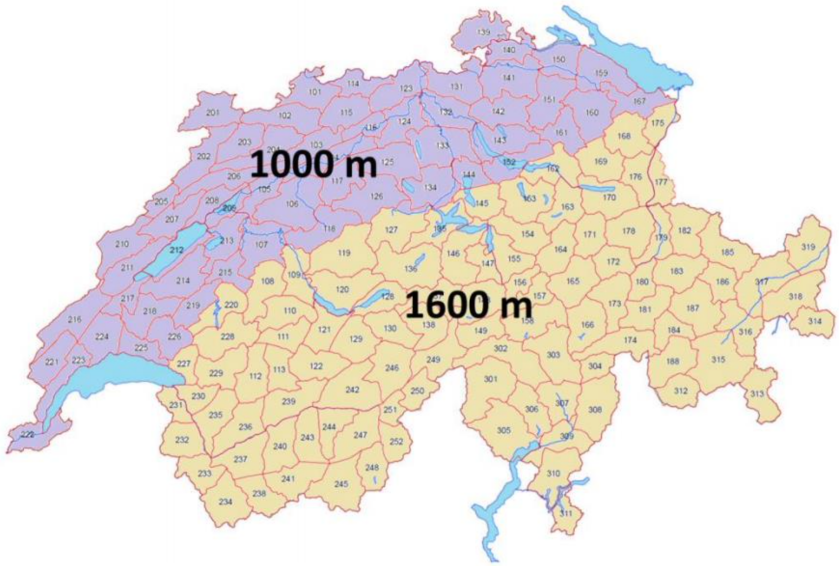


Figure 1.1: Warning map of Switzerland with colored height dependence (*Lowland/ Jura* blue, *Pre-Alps/ Alps, and Alp South Side* beige). The altitude thresholds are $1000\ mMSL$ and $1600\ mMSL$, respectively (c.f. Table 1.1). The smaller red regions depict the fixed warning regions (from (MeteoSchweiz 2021), pg. 4, Picture 3).

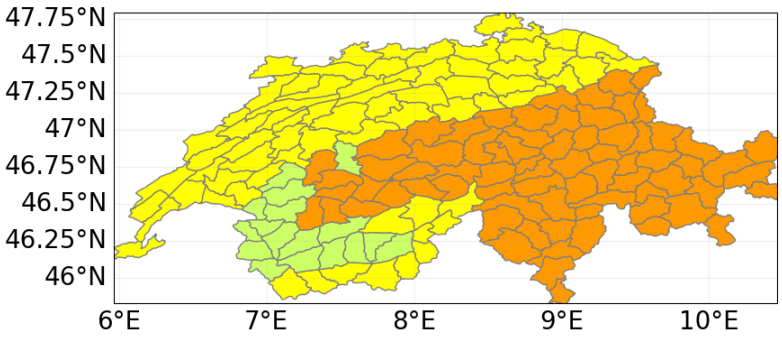


Figure 1.2: Example of a warning issued by forecasters. Warning for levels 1, 2, and 3 are issued for the fixed warning regions (depicted in grey). Example of a storm on October 30th 2018.

When a winter storm is approaching, forecasters evaluate the weather situation in the days ahead of the storm and can issue a warning. This warning can be refined by the forecasters up to the occurrence of the event. Forecasters can issue multiple warning levels for a warning region. The set levels are height specific, e.g., below $1800\ mMSL$ warning level 2 and above 4. In this example and as seen in Table 1.1, for a region in *Lowland/ Jura* from $1000 - 1800\ mMSL$ maximum wind gusts of $100 - 130\ km/h$ and from $1800\ mMSL$ onward maximum wind gusts of $160 - 200\ km/h$ are expected.

To analyze the quality of warnings, the objectives for warnings issued by MeteoSwiss are defined together with the Swiss Government (W. MeteoSchweiz 2017). Multiple scores are defined which compare the issued warnings to the maximum wind gusts measured during the event at fixed locations. A minimum threshold for each score is defined as a measure of quality.

The advantage of the current warning system is that weather forecasters can use NWP in combination with their expertise to generate a warning. Additionally, it is flexible as it allows the communication of different warning levels for specific heights, and forecasters can change the warning with the event coming closer (NWP and weather forecasts get more accurate closer to the event). Working with fixed warning regions reduces the complexity of the task and are easily recognizable by warning receivers.

The current system produces high-quality warnings allowing authorities and citizens to prepare for upcoming winter storms. However, there are currently two drawbacks which are addressed in this report. First, with fixed regions one loses spatial resolution of the warnings, which can lead to either too large, or too low area warnings. Using variable regions (warn polygons) would help mitigate this potential issue. Second, forecasters need to manually map a warning level to every region. With a partially automatized system, it would be possible to reduce the manual workload for forecasters allowing them to concentrate on more complex tasks for which their expertise is most needed.

1.3.2 Automatic Generation of Variable Warnings

In order to partially automatize the generation of warnings for winter storms, we developed an algorithm to produce and classify event-based warning regions based on the output of the COSMO2-E model. These event-based formed warning regions will here be termed *warning polygons*, and correspond to a set of points on the COSMO2-E grid which are assigned the same warning level. Throughout this report the warning levels and the associated thresholds of the current warning system (c.f., Table 1.1) will be used.

The above-mentioned goal can be divided into two hypotheses. (1) Warning polygons can automatically be generated from the COSMO2-E model outputs, are of sensible sizes, and can depict the danger levels from the maximum wind gusts of a winter storm. (2) Warning polygons can be defined as a collection of points on the COSMO2-E grid and are independent of the fixed warning regions of the current warning system (c.f., Figure 1.1).

1.4 Historical Windstorm Data

To test, fine-tune, and analyze the new warning generation method data of past winter storms will be used. MeteoSwiss archives all data needed over four years: the simulated maximum wind gusts from COSMO2-E, the measured maximum wind gusts, and the issued warnings. From 2018 to 2021, about 30 winter storms occurred in Switzerland which are used in the scope of this report. In addition, the building-up phases of winter storms as well as several non-stormy situations are added to the data set to include non-hazardous weather events. For simplicity, we shall consider one forecast per day, i.e., each data-point represents one day. The total data set contains 160 samples (days). These samples are divided into a training and a testing data set (ratio $7/3$). The training data set is used to design and fine-tune the algorithm. The test data set is used for testing only. In addition, due to the relative small amount of available data (since winter storms are extreme weather events by definition), the full statistical analysis described in Section 3.2 is done on the full data set.

As a first quality check, it would be sensible to compare the automatically generated warnings to the warnings issued by the forecasters. However, the here presented method

does currently only provide one warning per grid-point. But, as discussed in Section 1.3.1, forecasters can issue two distinct altitude-dependent warnings for the same region. Furthermore, this information is stored as explicit text in the MeteoSwiss archive, which renders an automatic extraction unpractical. Hence, a statistically robust comparison to the forecasters' warning was not achievable, and only qualitative comparison where conducted, in particular via a feedback-discussion with a few experts from MeteoSwiss.

As a second check, one can compare the warnings generated based on forecast outputs to the actual measured maximum wind gust values. MeteoSwiss measures maximum wind gusts at 300 to 600 locations in Switzerland shown in Figure 1.3. The weather stations are distributed all over Switzerland and located in cities as well as mountain tops. In this report, the wind speed levels (same levels as warning levels, see Table 1.1) measured throughout the day at every weather station are of interest. An example on October 30th 2018 is shown in Figure 1.3. Note that the weather stations' locations do not all align exactly with the COSMO2-E grid used for the warning generations, and thus the nearest value is always used for comparison.

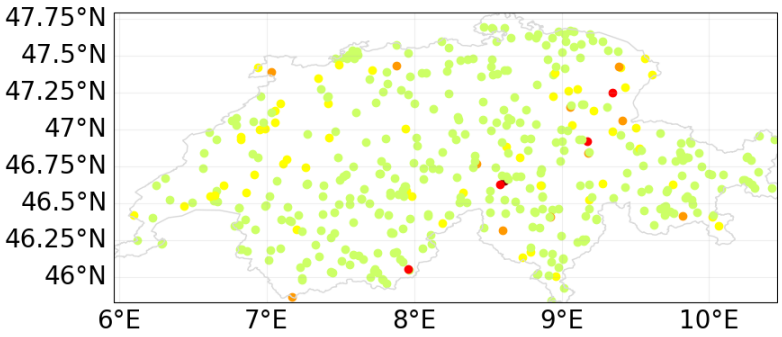


Figure 1.3: Maximum wind speed levels measured on October 30th 2018 at MeteoSwiss weather stations.

1.5 Related Work

Communicating complex NWP to weather forecasters and the public is a well-known problem in the weather forecasting community. As the high-performance computation capacities increased, models were able to forecast more weather variables which poses new challenges in visualization of model outputs (Roberts et al. 2019). Condensing and representing the modelled data in an appropriate way for weather forecasters is necessary to make use of the increased predictive value of NWP data (Roberts et al. 2019).

One option to present NWP is to visualize all individual ensembles of the NWP (e.g., 21 ensembles in COSMO2-E). The *Spaghetti Diagram* is an example of plotting all possible outcomes of a NWP and it shows the distribution of ensembles (Toth et al. 1997). This diagram can result in heterogeneous visualizations (Roberts et al. 2019) which are hard to interpret. The heterogeneity is precisely one of the main difficulties addressed with the new method presented in this technical report.

Post-processing methods, which reduce the heterogeneity in NWP, exist, e.g., neigh-

borhood approaches. In neighborhood approaches, for every grid point a spatial neighborhood is formed. In this neighborhood smoothing computations are performed which reduce the heterogeneity. For example, all grid points of the neighborhood can be considered as ensembles (*fractional coverage* (Blake et al. 2018)) or the probability that a set parameter threshold is exceeded in the defined neighborhood can be computed (*neighborhood maximum ensemble probability (NMEP)*) (Schwartz and Sobash 2017). NMEP was used to post-process updraft helicity (a diagnostic parameter that identifies rotating thunderstorm updrafts in storms (e.g., supercells)) by (Kain et al. 2008). In the NWP of the updraft helicity, a model resolution of $3 \times 3\ km$ and a neighborhood of $80 \times 80\ km$ was selected (Roberts et al. 2019). The result (see Figure 1c in (Roberts et al. 2019)) is still showing heterogeneity, despite the large neighborhood selected and a Gaussian smoother needed to be applied. Concluding, the NMEP and other state of the art methods cannot reduce the heterogeneity in NWP satisfyingly.

Therefore, a new approach to reduce heterogeneity in NWP is introduced in this report. The post-processing technique developed is a neighborhood approach. The algorithm developed for warning generation makes use of image processing techniques designed to reduce heterogeneity in images. This characteristic is used to reduce heterogeneity in NWP.

Chapter 2

Methodology

As initial inputs we consider the 21 forecast ensembles of maximum wind gust hourly-forecasts computed by the COSMO2-E model. The maximum wind gusts are first reduced to a single map per day by using a series of preprocessing steps explained in Section 2.1. Then, warning polygons are generated using simple but powerful image processing as described in Section 2.2. Finally, the quality of the new method is assessed by computing various scores such as the false alarm rate introduced in Section 2.3 and by comparison to the current warning system in Section 2.4.

2.1 Weather Data Preprocessing

Preprocessing of the COSMO2-E maximum wind gust outputs reduces their temporal dimension and number of ensemble members and results in a single map of points per day which is then used for the warning generation. This step while necessary, is not part of the core developments presented in this report. Future improvements in the weather forecasting model and in the preprocessing method can thus seamlessly be integrated in the here presented methodology.

2.1.1 Overview

The here used preprocessing entails four main steps. The initial data consist of the maximum wind gust values c on each grid point for each of the 21 ensembles and for each time (in general hourly values). First, the values are reduced to those in the time-window $[t_s, t_e]$. Second, the maximum value at each grid point is selected for each ensemble member separately. Third, for each grid point a single element corresponding to a fixed quantile t_g is selected from the ensembles. Fourth, the values are classified in the warn levels t_w (c.f. Table 1.1). The final output of the preprocessing is

$$p = f_t(f_g(f_c(c;\ t_s,\ t_e);\ t_g);\ t_w), \tag{2.1}$$

where

1. function f_c enforces the temporal constraint with start t_s and end time t_e,
2. function f_g groups ensemble members to obtain a single windfield map with parameter t_g (quantile, as explained in Subsection 2.1.3),

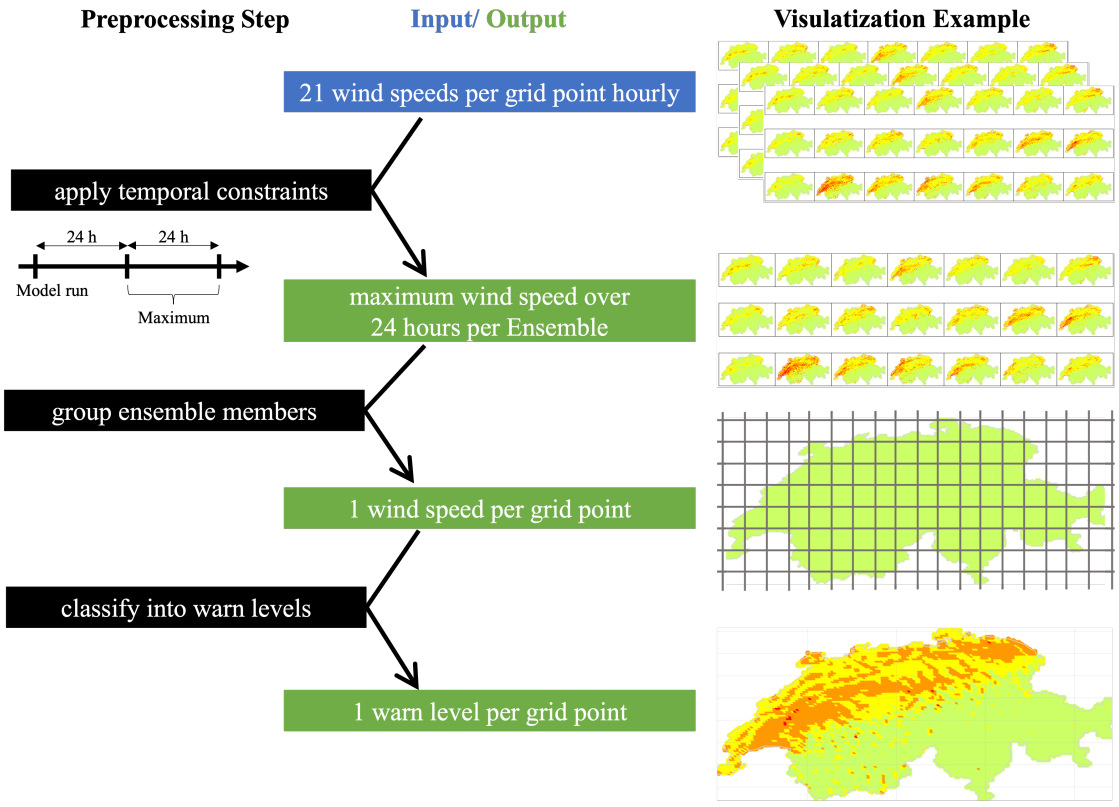


Figure 2.1: The weather data preprocessing steps (left) with their respective inputs and outputs (middle) and visual examples thereof (right) are shown.

3. function f_t classifies the grouped wind speeds into warning levels t_w (given in Table 1.1).

The preprocessing steps are illustrated in Figure 2.1. The structure of the data preprocessing follows the current procedure of handling the COSMO2-E output of MeteoSwiss. Note that variation of the preprocessing parameters (t_s, t_e, t_g, t_w) are however not part of this report, and it is expected that the presented algorithm to generate warm maps would be applicable to any parameter choice.

2.1.2 Temporal Constraint

The function f_c applies a temporal constraint on the COSMO2-E data output. As explained in Subsection 1.2, the COSMO2-E model forecasts weather up to five days ahead with an hourly resolution. For this report, the constraint is set to select maximum wind gust values between $t_s = 24$ hours up to $t_e = 48$ hours after the forecast beginning, as illustrated in Figure 2.1. The maximum wind gust per ensemble member and grid point over that time interval is then selected. The constraints could be adapted, e.g., by changing the start and end times, or by taking the average instead of the maximum value.

2.1.3 Grouping Ensemble Members

To reduce the 21 ensemble members per grid point to one wind speed per grid point, a statistical method f_g is applied. MeteoSwiss uses the 0.7^{th} quantile over the 21 wind speeds outputs for forecasts less than 36 hours away and the 0.4^{th} quantile for more than 36 hours. In this report, the 0.7^{th} quantile is taken ($t_g = 0.7$).

2.1.4 Warning Levels per Grid Point

The wind gust value per grid point are each classified into the given warning levels in Table 1.1 (t_w) by the function f_t. In this step, the fixed warning regions, shown in Figure 1.1, are used to map the boundary between *Lowland/ Jura* and *Pre-Alps/ Alps, and Alp South Side* to the COSMO2-E grid. The result of this step is the output p of the weather data preprocessing that consists in one warning level (value between 1-5) per grid point. The output possesses heterogeneous characteristics (see example in Figure 2.1), i.e., often a large difference in warning levels between neighboring points exists. Note that to avoid problems with boundaries, we consider a square boundary area of 15 grid points around the geographical extremal points of Switzerland (c.f., Figure 2.1).

2.2 Algorithm to Generate Warning Polygons

The heterogeneity in the warning levels coming directly from the processed COSMO2-E maximum wind gust data is hard to interpret for the people and authorities of Switzerland, because it is unclear whether the high-resolution corresponds to the actual precision of the forecast. It is thus hard to evaluate what really is the risk at a given location. The main contribution of this report is an algorithm designed to reduce the heterogeneity in the preprocessed data to produce warning polygons. The produced polygons must neither be too larger nor too small. In addition, the produced polygons should show some robustness to the uncertainty in the weather forecast. This is crucial so that the algorithm can be used to partially automatize the warning generation to support the weather forecasters.

The chosen algorithm consists of a series of filtering operations (convolutions). The core operations are known as morphological filtering, followed by a median filtering and a removal of too small regions. Generally speaking, filtering techniques have their origin in the field of image processing. They only use the information given in an image (in our case the heterogeneous map of warning levels). Thus, the procedure is not based on physical or meteorological knowledge. This has the advantage that any improvements in the forecasting model or in the preprocessing of the forecasting data (e.g., a more advanced treatment of ensemble forecasting uncertainty) will be directly and without further changes to the algorithm reflected in the final result.

2.2.1 Filtering

Morphological Filtering

Morphological filtering is based on the application of a succession of convolution operations of a predefined filter (shape to probe an image at every location and aggregate the values within the probe) on a given *binary* image. For each convolution, the filter is applied to the full image by translating the filter one point at the time. There are two

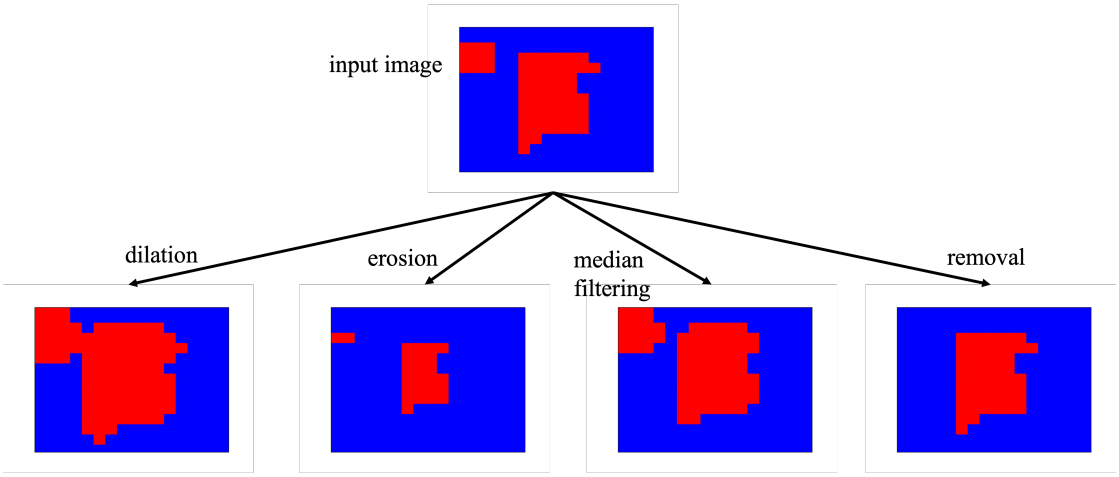


Figure 2.2: Example of technical operations (dilation, erosion, median filtering, and removal of too small regions) which are used in the warning algorithm. The area of interest is red.

main operations in morphological filtering: dilation and erosion, as illustrated in Figure 2.2 (c.f. also Equations 2.48, 2.49 in (Jain et al. 1995) and implementation by *scikit-learn* (Pedregosa et al. 2011)).

Dilation expands a region (red points in Figure 2.2) of points in a binary image by including every point to the region if one or more points of the binary image corresponds with the filter of size t_f (Jain et al. 1995). The application of the filter can be expressed as

$$A \oplus B = \bigcup_{b \in B} A_b, \tag{2.2}$$

where A is the binary image matrix, B is the filter, and b are all points within the filter. The size of the filter is a free parameter to be tuned to the specific task, in our case making warn polygons for wind gust forecasts. A larger filter size results in more expanded regions of interest (more red area in Figure 2.2), and smaller in less expanded regions of interest.

Erosion is the opposite operation of dilation and shrinks a region. Erosion only keeps a point in the specific region if every point of the filter B of size t_f corresponds with the binary image A (Jain et al. 1995). The application of the filter can be expressed as

$$A \ominus B = \{p |\ B_p \subseteq A\}. \tag{2.3}$$

If the filter is larger, the region of interest is shrunken more (less red area in Figure 2.2).

Median Filtering

The median filter is an averaging operation that smooths out an image, without blurring sharp discontinuities (Jain et al. 1995). It is based on a convolution and during the translation of the filter B of size t_s over a data image A, the median of the neighborhood is computed as

$$A \otimes B = \underset{b \in B}{\text{median}}\{A_b\}. \tag{2.4}$$

When this is applied to a binary image, it corresponds to a majority vote (see Figure 2.2). With a larger filter, the image is smoothed stronger, and more details get lost.

Removing too Small Regions

By applying the *Spaghetti Labeling* (c.f. (Bolelli et al. 2019) and implementation by *OpenCV* (Bradski 2000)) to detect connected components, multiples points forming one connected area can be found. The size of these areas can be computed and if they are smaller than a set threshold t_r they are removed (see Figure 2.2).

2.2.2 Algorithm's Design

The warn region generation algorithm consist in the repeated application of the previously described filtering operations (c.f. Section 2.2.1) starting from the warn map p obtained from the preprocessing of the COSMO2-E data (c.f., Section 2.1). For each warning level, starting from the highest one and up to the second-lowest one, the sequence dilation-erosion-dilation-smoothing is applied. All remaining non-classified points are set to the lowest level and the small regions are removed. For each of the warning levels (2-5), the parameters of the filters for the two dilations, the erosion and the smoothing filtering were calibrated using the training set of past winter storms as described in Section 2.2.2 and summarized in Table 2.1.

The algorithm, illustrated in Figure 2.3, can be summarized as follows:

1. Starting from the heterogeneous preprocessed warning level map, make a binary map with 0 in all points with the lowest warn level (1), and 1 for the highest warn level (5).
2. Apply the filtering operations dilation, erosion, dilation and smoothing one after another to the whole map.
3. Set to 1 all points from the preprocessed map of the highest remaining warn level that has not yet been used.
4. Repeat steps 2 and 3 until all but the lowest warn level have been processed as 1's.
5. Remove the too small regions.

A detailed implementation can be found in Appendix A. Further, the algorithm is implemented similarly in CLIMADA (Kropf et al. 2022) and is accessible for all.

Calibrating the Algorithm's Parameters

The filter sizes t_f and t_s for the morphological filtering and the smoothing are selected by a grid search. The filters for morphological filter have the shape of a disk (as used in image processing (Wilkinson and Roerdink 2009)) and for median filtering a rectangular shape (according to (Jain et al. 1995)). The options for these parameters, over which the grid search is done, are selected according to the experience of the algorithm's design. For the first dilation this corresponds to $[0, 1, 2]$ (small to only connect singular points), the erosion $[2, 3, 4]$ (medium to reduce too small generated areas), the second dilation $[6, 7, 8, 9]$ (large to generate polygons), and smoothing $[11, 13, 15]$ (large to reduce heterogeneity and generate visually smooth polygons). For the optimization on the training data set (c.f., Section 1.4), the score *Symmetric Extremal Dependence Index*, explained in Subsection 2.3.2, was used because it evaluates all quality aspects of a

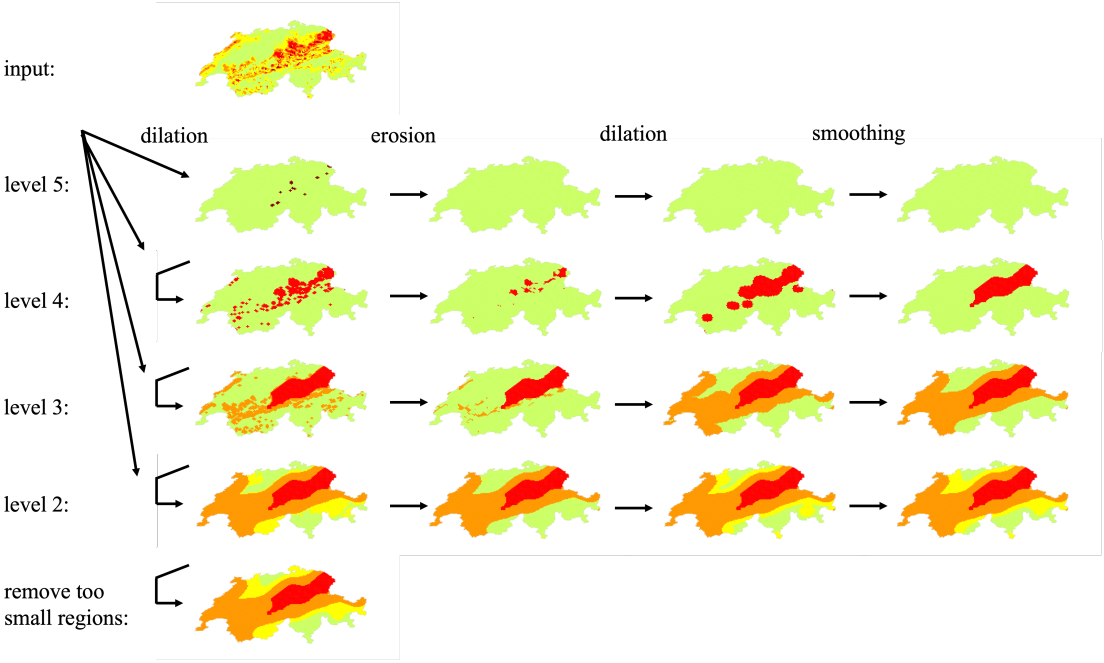


Figure 2.3: Algorithm 1 visually depicted for example winter storm (date: 20 Dec. 2019), where the formation of warning polygons is observed. Input (output of data preprocessing p) and output of every operation are shown.

	1. Dilation t_f	Erosion t_f	2. Dilation t_f	Smoothing t_s
Level 5	1	3	7	15
Level 4	1	3	7	15
Level 3	2	3	7	15
Level 2	1	4	8	15

Table 2.1: Parameter selection t_f and t_s for every warning level and functions f_f and f_s of the algorithm.

warning (hits, misses, false alarms, correct rejects, c.f. Table 2.2). The obtained optimal parameters are summarized in Table 2.1.

The parameter t_r for the removal of too small warning regions is selected based on the forecasters' expertise. After consultation with expert weather forecasters from MeteoSwiss, we consider warning polygons below $2000\ km^2$ as too small. This results in

$$t_r = \frac{2000\ km^2}{2.1\ km * 2.1\ km} \approx 450\ points, \tag{2.5}$$

where the $2.1\ km$ correspond to the resolution of the COSMO2-E grid used in this report.

2.3 Scores

In order to evaluate the performance of the warning generation algorithm, three different scores highlighting different aspects of a warning will be used. This is necessary because the answer to whether a warning is good or not is non-trivial. For instance, having too many false alarms creates the risk of a crying wolf, while missing a warning for a strong event can have devastating consequences. Since it was not possible to use the historical

warnings emitted by the forecasters as discussed in Section 1.3.1, we compare to the wind gust measurements at all available weather stations. More precisely, the maximum wind gust measurements per day per weather station are classified into the levels defined in Table 1.1. All the scores can then be derived from the contingency Table 2.2. Note that since the weather station location do not align with the COSMO2-E grid, the forecasted warning level of the geographically closest grid point to the weather station is taken (form pairs of forecasted warning level and measured level).

Contingency Table		**measurement**	
		yes	**no**
warning	**yes**	*a (hit)*	*b (false alarm)*
	no	*c (miss)*	*d (correct reject)*

Table 2.2: Contingency table to classify pairs of forecasted warning level and measured level. Can be computed for a single warning level or for all simultaneously. There are hits *a* (forecasted equals measured level), false alarms *b* (forecasted larger than measured level), misses *c* (forecasted lower than measured level), and correct rejects *d* (forecasted and measured level are 1).

The contingency table 2.2 can be derived for an individual warning level, or for all warning levels simultaneously. The former offers a separate analysis of every level of a warning, and the latter analyzes a warning over all warning levels. Generally, to fill the contingency table all points where a measured warning is defined, we extract the geographically corresponding point in the forecasted warnings (COSMO2-E or algorithm). These pairs of points are then classified in hits *a* (forecasted equals measured level), false alarms *b* (forecasted larger than measured level), misses *c* (forecasted lower than measured level), and correct rejects *d* (forecasted and measured level are 1).

If a warning level is analyzed individually, all pairs of points for which neither the forecasted nor the measured warning are of the desired level are counted as correct rejects d. The rest of the contingency table is filled using the remaining pairs. For instance, suppose we analyze warning level 3.

- Forecasted warning level 3 and measured level 3 is counted as hit a.
- Forecasted level of 3 and measured level of 1 or 2 is counted as false alarm b.
- Forecasted level of 4 or 5 and measured level of 3 is counted as false alarm b.
- Forecasted level of 1 or 2 and measured level of 3 is counted as miss c.
- Forecasted level of 3 and measured level of 4 or 5 is counted as miss c.
- Forecasted level of 1, 2, 4, or 5 and measured level of 1, 2, 4, or 5 is counted as correct reject d.

Note that MeteoSwiss adapted the computation of the contingency Table 2.2, e.g., by adding a tolerance to the wind gust measurements (see (W. MeteoSchweiz 2017)). Therefore, the computed scores of MeteoSwiss are not directly comparable to the results computed by filling the contingency table.

All the used scores can be expressed as

$$S = s(v_m,\ w,\ t_i,\ t_{d_v},\ t_{d_h}), \tag{2.6}$$

with the measured warning levels v_m and the generated warning levels w which are used to fill the contingency Table 2.2, the warning level of interest t_i, the maximal vertical t_{d_v} (set to $500\ m$) and horizontal t_{d_h} (set to $200\ m$) distances used for weather station to grid point nearest-neighbour matching.

2.3.1 Probability of Detection and False Alarm Ratio

MeteoSwiss reports the scores *Probability of Detection (POD)* and *False Alarm Ratio (FAR)* to the Swiss Government.

The POD is the conditional probability $\mathbb{P}$ that a generated warning is of level t_i given that the measured warning level is t_i, and can be approximated (c.f. Equation 2 in (McBride and Ebert 2000)) as

$$POD = \mathbb{P}(w = t_i |\ v_m = t_i) = \frac{a}{a+c}, \tag{2.7}$$

with a the number of hits, and c the number misses. If no hits a and misses c are detected ($a + c = 0$), the score POD is not defined. POD achieves the optimal value 1 if no misses and at least one hit are counted. This score includes the number of hits and the number of misses in a warning but ignores the false alarms and correct rejects. Therefore, POD rates a warning only on its skill of forecasting a measurement correctly. For example, if one measurement out of 400 is of level 3 and the rest of level 1, warning level 3 for all of Switzerland would achieve a perfect POD. This is called hedging, using the scores weaknesses to achieve a good score independent of the true warning quality.

The FAR is the conditional probability that the measured warning is different from t_i given that the algorithm generated warning level is t_i. It can be approximated (c.f. Equation 3 in (McBride and Ebert 2000))

$$FAR = \mathbb{P}(v_m \neq t_i |\ w = t_i) = \frac{b}{a+b}, \tag{2.8}$$

with a the number of hits and b the number false alarms. If no hits a and false alarms b are detected ($a + b = 0$), the score FAR is not defined. FAR is optimal (value 0) if no false alarms and at least one hit are counted. Hedging is restricted if a warning system is optimized to perform well in POD and FAR simultaneously.

2.3.2 Symmetric Extremal Dependence Index

The score *Symmetric Extremal Dependence Index (SEDI)* (c.f. Equation 2 and Definition of H, F in (Ferro and Stephenson 2011)) does not allow hedging by definition, because all variables (a, b, c, d) of the contingency Table 2.2 are included. The score can be expressed as

$$SEDI = \frac{\ln(f) - \ln(h) + \ln(1-h) - \ln(1-f)}{\ln(f) + \ln(h) + \ln(1-h) + \ln(1-f)}, \tag{2.9}$$

where SEDI corresponds to the calculated score and

$$\begin{aligned} h &= \mathbb{P}(w = t_i |\ v_m = t_i) = \frac{a}{a+c}, \\ f &= \mathbb{P}(w = t_i |\ v_m \neq t_i) = \frac{b}{b+d}, \end{aligned} \tag{2.10}$$

where h corresponds to the POD and f to the FAR. For the score SEDI to be defined, no entry of the contingency table can be zero, else h or f are zero or one, respectively, and the natural logarithm of zero is not defined (Equation 2.9 includes e.g., $ln(f)$ and $ln(1-f)$). The computed value of SEDI lies in $[-1,\ 1]$ because, unlike in POD and FAR, the numerator can be negative for large f and small h. If the computed score is less or equal to zero, the computed warning is said to have no skill in forecasting the event. This is justified because a warning with a negative score must comprise many false alarms and misses which result in a low h and high f. The SEDI is maximized (optimal value 1) if $h \to 1$ (perfect hits) and $f \to 0$ (zero misses). The SEDI is non-degenerate for rare events, i.e., it does not tend to zero for small base rates ($\mathbb{P}(v_m = i) \to 0$) because it only depends on h and f (Ferro and Stephenson 2011). SEDI is base rate independent and suited for rare events (Ferro and Stephenson 2011), which is an advantage compared to the POD and FAR.

2.4 Differences in Warning Generations

The current warning system used by MeteoSwiss requires forecasters to map warning levels onto fixed warning regions. These regions are analogous to the algorithmically generated warning polygons that consist of a non-fixed collections of COSMO2-E grid points. The weather forecasters make use of their extensive passive knowledge of weather patterns and the COSMO2-E strengths and weaknesses. On the contrary, the algorithmically warning generation relies entirely on the COSMO2-E maximum wind gust output and cannot compensate for its forecasting errors directly. Thus, it is to be expected that the automatic warning generation cannot perform as well as the forecasters. As the historic warning data was not directly obtainable from the MeteoSwiss archive, we were not able to make a direct quantitative comparison in this report (c.f. the discussion in Section 1.3.1). Instead, we compare the algorithm's scores with respect to the measured maximum wind gusts to the scores obtained from the heterogeneous output of the COSMO2-E model (the output of the preprocessing state). It is clear that the algorithm cannot be more skillful in the warnings. However, as shown in Section 3.1, the algorithm warning generation can achieve comparable scores while generating warning polygons that were judged to be of sufficient quality (visually) by expert forecasters.

Chapter 3

Results

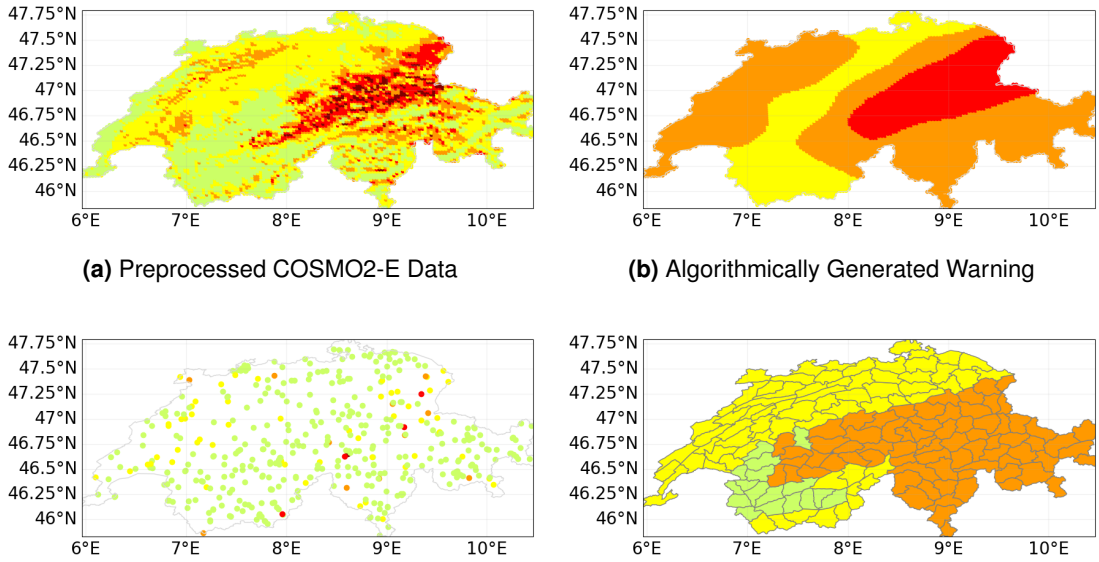
3.1 Automatically Generated Winter Storm Warnings

In this section, three examples of past winter storms are described. The scores for the automatically generated warning polygons are reported, but no statistical comparison to the preprocessed direct COSMO2-E output is done here (see Section 3.2 for a full statistical comparison). The first storm occurred on October 30th 2018. For this storm, the final warning map issued by the forecasters is available for comparison. This storm was overestimated by the COSMO2-E model, which was identified by the forecaster who thus issued a lower warning. The second storm occurred on December 14th 2019 and was a rather weak storm. The third storm that happened on February 4th 2020 was quite strong, and affected most of Switzerland.

3.1.1 Winter Storm on 30 Oct. 2018

For the winter storm occurred on October 30th 2018, most classified maximum wind gust measurement data (measurements classified into levels of Table 1.1) lie in level 1, as shown in Figure 3.1c. At few weather stations maximum wind gusts of level 2, 3, 4, and 5 were measured. Most higher values were measured in the East and South of Switzerland. The preprocessed COSMO2-E data shown in Figure 3.1a forecasted a large area with maximum wind gust of levels 4 and 5 in the East, as well as level 3 in the West. Thus, the storm was quite significantly overestimated by the model.

The overestimation by the COSMO2-E model translates directly into the algorithmically generated warnings shown in Figure 3.1b. A warning polygon of level 4 is produced in the East, surrounded by a large level 3 warning polygon. Thus, the few measurements of level 4 in the East are warned correctly, but many false alarms occur since several warning level 1 wind gusts measured in the East and South East were algorithmically labelled to be of level 3 and 4. The level 3 warning polygon in the North West of Switzerland coincides with two measured levels only and the rest are false alarms. Therefore, this warning polygon, which is based on the preprocessed COSMO2-E data, could be considered unnecessary. Note however the sparse density of weather stations in the North West. Similarly, the warning polygon of level 2 in Central Switzerland is mostly incorrectly attributed by the algorithm as the measurements show it to be of level 1. These unnecessary warnings result in a FAR of 0.97 of the algorithmically generated warning. Since the warning polygons are at most weather station locations of higher



(a) Preprocessed COSMO2-E Data

(b) Algorithmically Generated Warning

(c) Classified Measurement Data

(d) Forecaster's Warning

Figure 3.1: Winter storm that occurred on October 30th 2018. The preprocessed COSMO2-E data (a) overestimate the situation compared to the maximum wind speed levels measured (c). Therefore, the algorithmically generated warning (b) produces too large warning polygons and generates many false alarms. The forecaster's warning (d) did not issue a level 4 warning and reduced the false alarms.

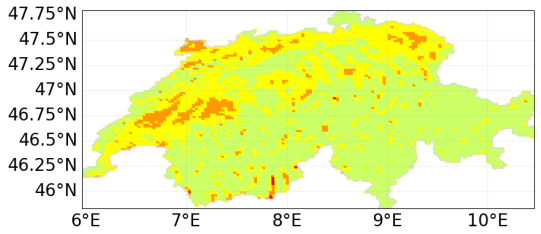
levels than the measured ones, the number of hits and misses are low. These result in POD of 0.5 of the algorithmically generated warning. Since there is no level 1 warning polygon, the correct rejects d entry of contingency Table 2.2 is zero, and therefore the score SEDI is not defined for this warning (c.f. Subsection 2.3.2).

As shown in Figure 3.1d, the forecasters did not issue a warning of level 4 and the areas warned in level 2 and 3 are considerably smaller compared to the algorithmically generated warning. However, we note that also the forecasters seem to have overestimated the storm in their warning. Most areas are warned with either level 2 or 3, while the measurements reported mostly level 1 values. In addition, none of the level 4 measurements was warned for. This might however be due to variations in the height, as it is not rare to have strong maximum wind gusts at high altitudes. The POD is 0.75, the FAR 0.91 and SEDI -0.25 of the warning issued by the forecasters. Therefore, as expected, the latter achieves higher POD and lower FAR scores than the algorithmically generated warning.

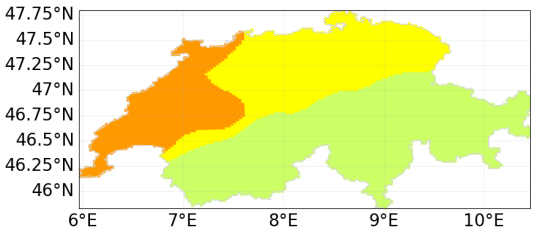
3.1.2 Winter Storm on 14 Dec. 2019

For the winter storm that occurred on December 14th 2019, most measured maximum wind gusts (c.f. Figure 3.2c) are of level 1 in the South, and South East of Switzerland. A band from the South West to the North East of level 3, 4, and 5 maximum wind gusts is measured. The preprocessed COSMO2-E data (see Figure 3.2a) did not forecast any level 4. The level 3 maximum wind gust are forecasted roughly in the same regions as eventually measured, but with a very heterogeneously distribution.

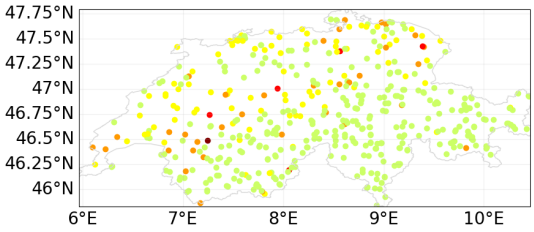
The algorithm removes the heterogeneity of the preprocessed COSMO2-E wind gust



(a) Preprocessed COSMO2-E Data



(b) Algorithmically Generated Warning



(c) Classified Measurement Data

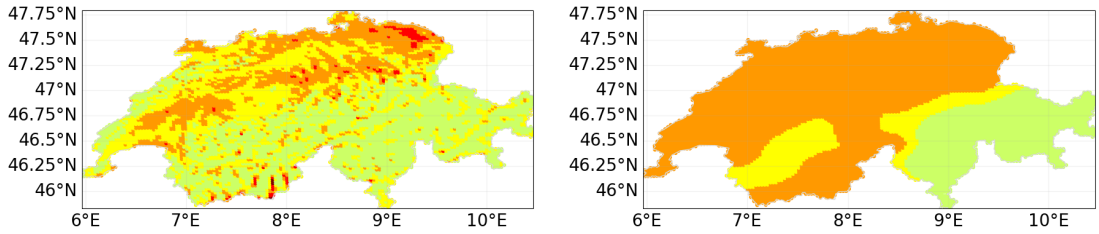
Figure 3.2: Winter storm that occurred on December 14th 2019. The COSMO2-E forecast (a) underestimates the storm compared to the measurements (c). The algorithmically generated warning polygon (b) of level 3 is too small, and therefore misses most measured maximum wind gusts of level 3 and 4.

data and produces a single warning polygon for each of the levels 1, 2, and 3, as shown in Figure 3.2b. Thus, all level 4 maximum wind gust measurements are missed. The warning level 3 polygon covers the North West of Switzerland and therefore misses the measured level 3 maximum wind gusts in the center and North East of Switzerland. The level 2 warning polygon cover the rest of the South West - North East band where maximum wind gusts above level 1 were measured. Overall, the automatically generated warning polygons results in a POD of 0.49 and an FAR of 0.7. In the South, the warning level 1 comprises many correct rejects which results in a SEDI score of 0.25 which indicates that the warning has skill in forecasting the winter storm. Note that the measurements are quite heterogeneous too, which can be attributed to strong differences in measured maximum wind gusts at different altitudes. The forecaster's warning for this day is accordingly height dependent. Due to the difficulty in extracting the height information from the archived data (see Section 2.4), the comparison between the algorithmically generated and forecaster's warning are not done in this report.

3.1.3 Winter Storm on 04 Feb. 2020

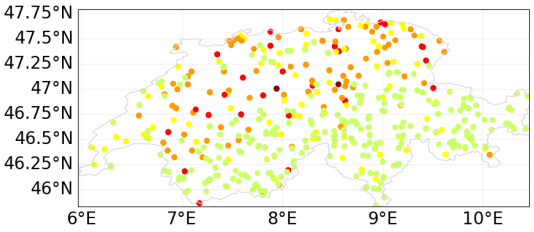
For the winter storm that occurred on February 4th 2020, many maximum wind gusts of level 3 and 4 are measured (see Figure 3.3c). These levels are mostly measured in the Center, North, North East, and West of Switzerland. The South and South East are dominated by level 1 measurements. The preprocessed COSMO2-E data in Figure 3.3a shows level 3 danger for the area the storm was measured in. The preprocessed COSMO2-E data also contains a few level 4 points in the North East and in the South West.

The automatic warnings generation results in one larger level 3 polygon covering the



(a) Preprocessed COSMO2-E Data

(b) Algorithmically Generated Warning



(c) Classified Measurement Data

Figure 3.3: Winter storm that occurred on February 4th 2020. The preprocessed COSMO2-E data (a) correspond well to the measured maximum wind speed levels (c). The warning polygon (b) of level 3 corresponds well with the measured level 3 maximum wind gusts but forecasting level 4 measurements is not achieved.

West and the North, as well as parts of the South West as shown Figure 3.3b. In addition, two level 2 polygons cover parts of the Central West and East. Finally, the South East is covered by a level 1 polygon. The heterogeneity has thus been successfully reduced. No level 4 or 5 warning polygons were generated, as the COSMO2-E level 4 points are too heterogeneously distributed. Thus, all level 4 and 5 measurements are missed, while a larger amount of level 3 warnings in the North are hits. Overall, we obtain a POD of 0.51 for the warning. The part of the warning polygon of level 3 in the South West of Switzerland produces mostly false alarms, as most measurements are of level 1. The linking of the northern and southern parts into one polygon of level 3 (c.f. Equation 2.5) is an artefact from the second dilation of the algorithm. Overall, the FAR is 0.72.

For large parts of the North level 4 were measured, which indicates that a level 4 warning would have been more appropriate. However, the measurements are heterogeneous, and intermixed with levels 2 and 3. Thus, a warning level 4 would have increased the POD at the cost of a higher FAR. The obtained level 3 polygon in the North results in a SEDI score of -0.26, which indicates no skill.

Similarly to the storm on December 14th 2019, the heterogeneity in the measured maximum wind gusts can be explained by altitude differences. Accordingly, the forecaster's warning for this day is height dependent. Hence, again no comparison was possible due to the difficulty in extracting the height information from the archived data (see Section 2.4).

3.2 Statistical Analysis

We will now analyze the robustness of the automatic warning generation algorithm compared to the output from the COSMO2-E model for the full set of storms from 2018 - 2021 (as described in Section 1.4). In order to increase the statistical significance, we consider the full set containing both the training and the testing data. This doesn't allow to conclusively asses the robustness of the algorithm quantitatively, but in combination with the feedback from the forecasters (c.f. Section 3.3) it allows to qualitatively discuss the suitability of the algorithm for semi-automatizing the warning pipeline.

For all the storms in the data set the scores POD, FAR and SEDI are computed for both the automatically generated warnings and the preprocessed output from the COSMO2-E model using the reported measured values at the weather stations (c.f. Section 1.4). The results are reported in the form of *Violin Plots* that represent the full distribution (colored area) of the scores over all samples, the median (white dot), the interquartile ranges (thick black line from first start to third quartile), and adjacent values (light black line).

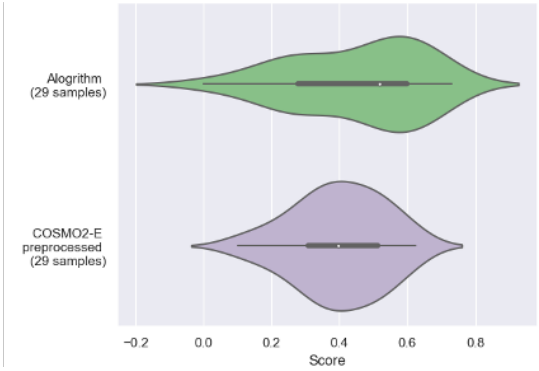
Since we consider warnings for extreme events, which by definition are rare, the data set is heavily skewed towards lower warning levels. Thus, we first consider each of the warning levels separately. The results for warning levels 4 and 5 are not reported due to the small number of storms occurring in these levels. Second, all warning levels simultaneously are considered. Samples for which a score is undefined are ignored. Thus, the total samples for a given score may differ between the algorithmically generated warnings and the COSMO2-E warnings.

3.2.1 Warning level 1

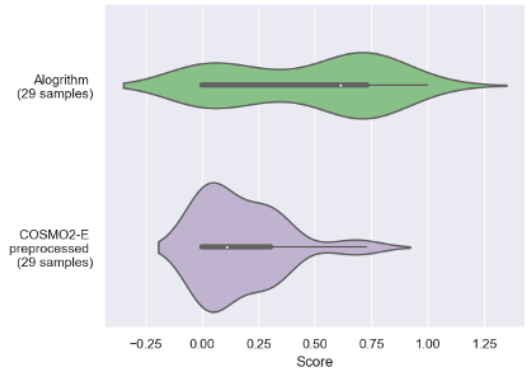
The POD has no informative value for warning level 1 because no misses can occur when analyzing warning level 1 only. The FAR median of the algorithm is 0.65 and the distribution of scores is small (see Figure 3.5a). The relatively large median is due to algorithms tendency to generated large warning polygons. The median of 0.72 of the COSMO2-E preprocessed data is even higher and is likely due to extreme heterogeneity of the forecasted warnings since we compare only to the closest measured value. The lower distribution of achieved FAR scores of the algorithm compared to the COSMO2-E preprocessed shows that the reduction in heterogeneity lowers the number of false alarms for level 1. For the SEDI score, shown in Figure 3.6a, the algorithm has median of 0.41 which is lower than the COSMO2-E preprocessed value of 0.44. The distribution of the algorithm shows a wider spread of values as the COSMO2-E and therefore performs not consistent in level 1. Despite the higher FAR, the COSMO2-E preprocessed has slightly more skill than the algorithm in level 1.

3.2.2 Warning level 2

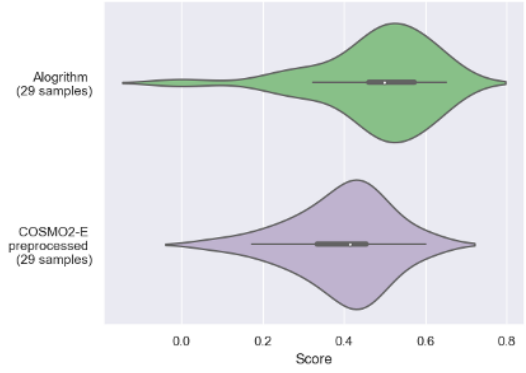
For warning level 2, the algorithm outperforms the COSMO2-E preprocessed warnings in every score. The median POD of the algorithm is 0.52 compared to 0.4 for the COSMO2-E preprocessed. However, the POD distribution (see Figure 3.4a) of the algorithm shows a lower adjacent value of zero, which means that the algorithm is inconsistent in level 2. The median in FAR achieved by the algorithm is 0.11 compared to 0.25 for the COSMO2-E preprocessed. The distribution of the FAR of the COSMO2-E preprocessed



(a) POD Warning level 2



(b) POD Warning level 3



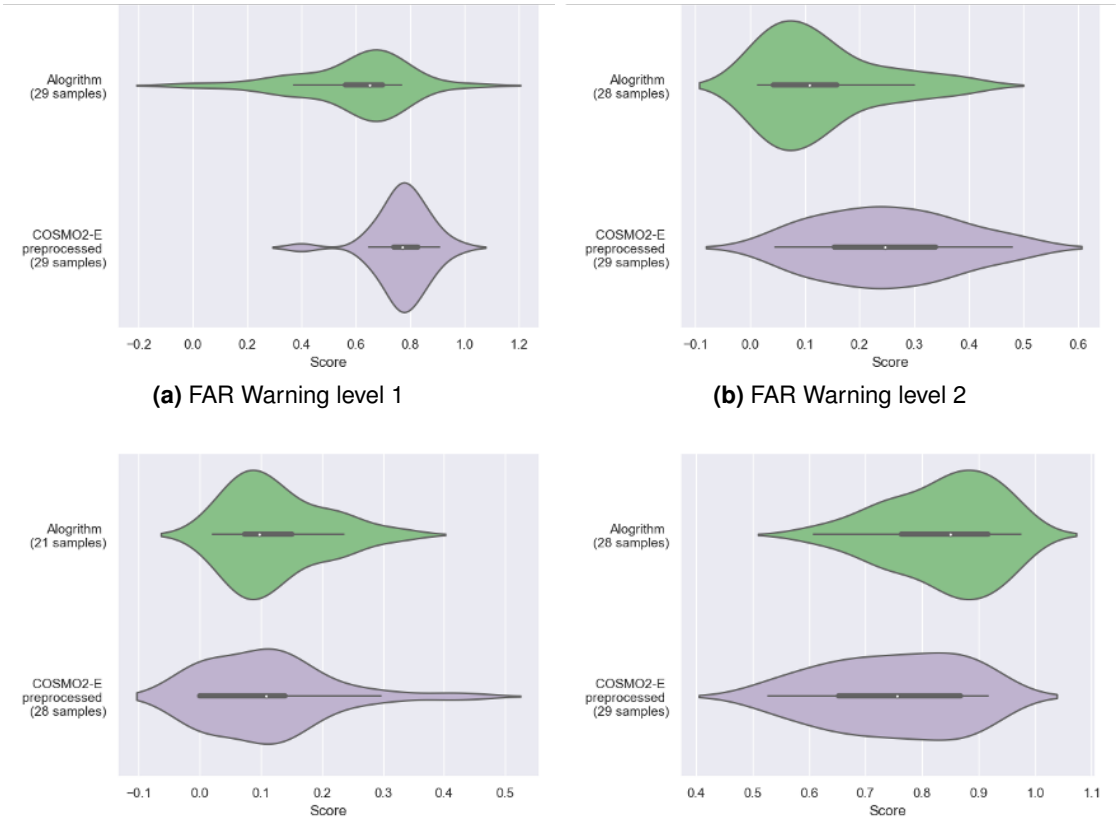
(c) POD over all Warning levels

Figure 3.4: Violin plots of probability of detection for warning levels 2, 3, and over all warning levels (level 1 no informative value, because no misses can occur for warning level 1). The number of samples shows how many days are used to compute the distribution of scores. The algorithm outperforms the COSMO2-E preprocessed in all statistics.

shows a large distribution. The POD and FAR of the COSMO2-E preprocessed are likely worse compared to the algorithm, because of its heterogeneity. The results of the POD and FAR are supported by SEDI but the difference in medians is smaller (algorithm 0.38 and COSMO2-E preprocessed 0.33) as seen in Figure 3.6b. The algorithm is not consistent in achieving scores, as can be seen in the large distribution of SEDI. Nevertheless, the analysis of SEDI underlines that warning level 2 profits from the reduction of heterogeneity and produced warning polygons.

3.2.3 Warning level 3

The algorithm also performs better in warning level 3 compared to the COSMO2-E preprocessed warnings. Forming warning polygons for warning level 3 increases the POD, from a median of 0.11 for the COSMO2-E preprocessed to 0.61 for the algorithm. However, the low first quartile of the algorithm's POD distribution indicates a wide spread of scored values, as shown in Figure 3.4b. The improvement in the POD by the algorithm does not influence the FAR (both median 0.1, see Figure 3.5c). The distribution of FAR scores is similar for both approaches. The heterogeneity in the COSMO2-E preprocessed produces many misses for level 3, which results in a low POD and does not change the FAR. The algorithm has a higher skill, with a SEDI score of 0.4 compared



(a) FAR Warning level 1

(b) FAR Warning level 2

(c) FAR Warning level 3

(d) FAR over all Warning levels

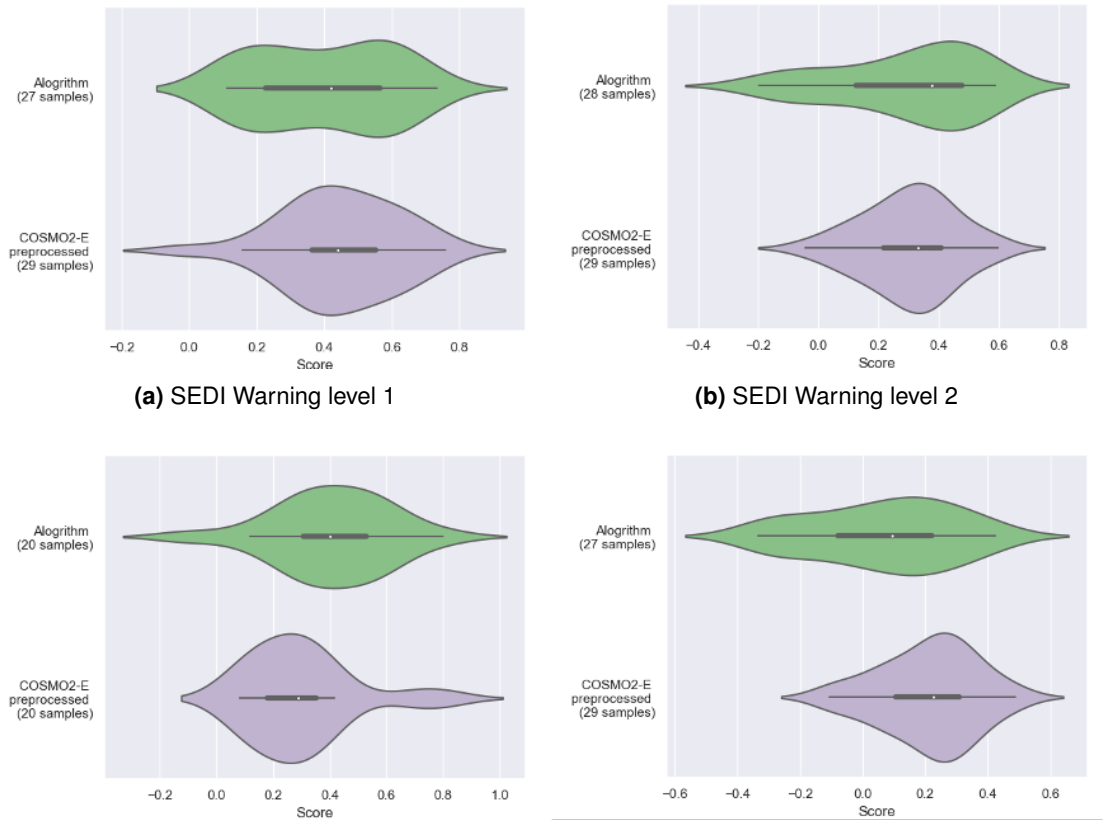
Figure 3.5: Violin plots of false alarm ratio for warning levels 1, 2, 3, and over all warning levels. The number of samples shows how many days are used to compute the distribution of scores. The COSMO2-E preprocessed has higher (worse) medians than the algorithm in separate analyses. However, over all warning levels the algorithm has a higher median.

to 0.29. The distribution of SEDI scores shows a wider spread of values for the algorithm; nevertheless, the lower adjacent value is similar.

3.2.4 Statistical Analysis over all Warning Levels

For an overall comparison of the automatically generated warnings and the COSMO2-E preprocessed outputs we compute the scores for all levels together, including levels 4 and 5. The absolute value of all scores are for both methods worse than for the individual warning levels as shown in Figure 3.6d. This is likely because for the single level analysis, while the hits are all included, not all misses are accounted for. Indeed, we recall that for a single level, all pair of points for which neither the forecast nor the measured levels are of the desired level are defined as correct rejects (c.f. Section 2.3).

In the POD (see Figure 3.4c) the algorithm with median 0.48 is favorable compared to the COSMO2-E preprocessed with median 0.39. This is in correspondence with the separate warning level analysis of POD. However, in the FAR the algorithm with a median of 0.85 performs worse overall than the COSMO2-E preprocessed with 0.75, as shown in Figure 3.5d. Clearly lower scores are computed for the COSMO2-E preprocessed than for the algorithm, as can be seen by the distribution of scores. This can be explained by the warning polygons of levels 2, 3, and 4 generated by the algorithm, which cover more



(a) SEDI Warning level 1

(b) SEDI Warning level 2

(c) SEDI Warning level 3

(d) SEDI over all Warning levels

Figure 3.6: Violin plots of symmetric extremal dependence index for warning levels 1, 2, 3, and over all warning levels. The number of samples shows how many days are used to compute the distribution of scores. The algorithm outperforms the COSMO2-E preprocessed in separate analyses. However, over all warning levels the COSMO2-E preprocessed has a higher median.

area than singular points in these warning levels of the COSMO2-E preprocessed. This is particularly evident in e.g., Figures 3.2a and 3.2b. Thus, the reduction in heterogeneity resulted in a higher POD, but also a higher FAR. Interestingly, the algorithm's higher FAR value over all warning levels leads to a different conclusion than in the separate analysis, were the FAR was lower. The lower values of the algorithm compared to the COSMO2-E preprocessed in the separate analysis were surprising, as an increase in false alarms caused by the reduction of heterogeneity from the generation of warning polygons was expected. It is not yet understood why this is only reflected in the analysis over all warning levels. It can be concluded from the POD and FAR over all warning levels that reducing the heterogeneity in the COSMO2-E preprocessed leads to a higher POD at the cost of a higher FAR. The score SEDI includes the false alarm rate and POD and therefore quantifies the trade-off shift of the algorithm towards a better POD and worse FAR. SEDI over all warning levels, shown in Figure 3.6d, infers that this trade-off leads to warnings generated by the algorithm with less skill in forecasting measured levels than the COSMO2-E preprocessed has. This is seen in the lower distribution of scores, and the median of 0.09 of the algorithm and the 0.23 of the COSMO2-E preprocessed. This large difference is surprising, because the difference in medians in the POD and FAR are similar and therefore an offset of these could have been expected. Why this is not the case could be, firstly, because of the false alarm rate being used in SEDI (and

not the false alarm ratio (FAR)) which uses the correct rejects of a warning. It is possible that the algorithm has fewer correct rejects and therefore a higher SEDI. Secondly, the false alarms could have a larger impact in SEDI. Since the algorithm has more false alarms, the lower SEDI median is the consequence.

3.3 Feedback Weather Forecasters

Examples of the algorithmically generated warnings, as well as the statistical analysis presented in Section 3.2, were presented to MeteoSwiss expert weather forecasters to receive feedback and improve the warning generation. The following key messages where distilled form these conversations:

- The warnings generated by the algorithm are an improvement compared to the previous automation attempt (AutoWARN). Especially, an improvement is achieved in a sensible reduction of the warning polygons heterogeneity.
- Small regions (links) connecting larger warning polygons to form a single larger warn polygon appear to be a meteorologically unjustified artefact of the warning generation algorithm (see e.g., the link of the level 2 polygon in the center of Switzerland in Figure 3.3b).
- As explained in Section 2.4, forecasters can issue height dependent warnings and the implemented algorithm cannot. The unavailable height information, explained in Section 1.4, hinders a direct comparison between the automatically generated warnings and the warnings in the MeteoSwiss archive.
- Forecasters regard warning polygons below $2000\ km^2$ area for contiguous polygons as too small. This information is used in Equation 2.5 to make sure that the algorithm does not produce smaller warning polygons.

Chapter 4

Discussion and Outlook

4.1 Discussion of Hypotheses

The goal of this report was to design and implement an algorithm which can automatically produce and classify event-based warning polygons for winter storms, based on the COSMO2-E maximum wind gust output. Three automatically generated warning examples in Section 3.1 show that event-based polygons can be formed automatically by an algorithm not relying on fixed warning regions. The proposed algorithm uses *Morphological Filtering* to reduce heterogeneity in the preprocessed COSMO2-E maximum wind gusts and produce warning polygons. The produced polygons were judged by expert forecasters to be a good basis for an automatic warning generation support system. However, the algorithm showed linkages between warning polygons which are considered artefacts by the expert weather forecasters. These are caused by the algorithm's design which connects smaller areas to larger polygons of the same warning level.

4.2 Accuracy of Algorithmically Generated Warnings

The most interesting statistical test of the algorithm would be to compare the algorithmically generated warnings to the warnings issued by weather forecasters. The algorithm is fully dependent on the COSMO2-E model outputs, whereas weather forecasters additionally use their expertise of winter storms when generating warnings. This allows them to correct for errors in the numerical forecast. For instance, for the winter storm on October 30th 2018, the issued warning achieved higher statistical scores than the algorithmically generated warning. Unfortunately, in this study we were not able to extract the height information of the forecaster's warning from the archived data. Thus, a direct comparison for all storms could not be done. Instead, we compared the algorithmically generated warnings and to the preprocessed maximum wind gust output from the COSMO2-E model. This quantifies the change in achieved scores resulting from the reduction in heterogeneity of the numerical forecast. It can thus be expected that the algorithm's warning will improve with each improvement of the numerical forecasting model.

We found that the algorithmically formed warning polygons have a higher chance of correctly assigning the measured level (higher POD) than the preprocessed COSMO2-E data. However, the formed polygons and the reduced heterogeneity in the algorithmically

generated warnings cause the over all FAR to increase. Therefore, shaping polygons and reducing heterogeneity in the COSMO2-E maximum wind gust decreases the number of missed forecasts at the cost of false alarms. These results were expected, because larger areas for levels 2, 3, and 4 increase the chance of forecasting the measured warning levels correctly. The trade-off between hits, false alarms, and misses was quantified by the SEDI score. The algorithm performs better than the COSMO2-E preprocessed when considering the warning levels 2 and 3 separately. However, considering all warning levels, the algorithm is outperformed by the COSMO2-E preprocessed data. The difference in median SEDI scores is unexpected by looking at the POD and FAR over all scores, and could be because the SEDI score penalizes false alarms more than misses. Taken together, reducing details in the COSMO2-E preprocessed maximum wind gusts to produce warning polygons impacts the scores differently (POD higher, FAR higher, and SEDI lower). Nevertheless, we can conclude that the scores achieved by the algorithm are comparable to the ones of COSMO2-E preprocessed wind gusts. Thus, the realized reduction in heterogeneity does not compromise the quality of the warning.

4.3 Towards Fully Automatic Warnings

At the beginning of this report, we developed two hypotheses which were not addressed and therefore remain open for future work. The first hypothesis, assigning an uncertainty to every warning level mapping to COSMO2-E grid point, could be approached by combining the unprocessed COSMO2-E data together with the algorithmically generated warning. This allows to take the distribution of ensemble members of the COSMO2-E model into account and combining it with the neighborhood of grid points which is considered by the algorithm. The second, generating a warning with a temporal constraint (warning start and end), could be approached by adapting the implemented algorithm to compute hourly preprocessed COSMO2-E data. If the latter exceeds a threshold, the starting point of warning can be set. If the threshold is not exceeded anymore, the end of the warning can be set. Over this time interval, a warning can be computed by developing our algorithm from 2D to 3D. The morphological filtering operations therefore need to be 3D which can be achieved by defining 3D filters (cylinders for morphological filtering and cuboid for median filtering). The filters are then applied to combine all hourly warnings in the defined time interval and produce warning polygons which hold over the defined time interval.

We decided to implement an algorithm based on the image processing techniques and were encouraged by the achieved results and the weather forecasters' feedback. We could have also decided to implement an approach based on machine learning, e.g., convolutional neural networks. We decided against this due to the small available data set, a missing definition of the perfect warning (imperfect scores of Subsection 2.3), and confidence in the filtering approach. Nevertheless, the learning-based approach could still be implemented in future.

In a broader perspective, given the positive feedback from the expert warning forecasters, and given the comparable scores of the warnings generated by the proposed algorithm and of the warnings obtained directly from the COSMO2-E outputs, we conclude that the presented method is a promising step in the building of a partially automatic weather warning system. Furthermore, the proposed algorithm could be tuned and directly applied to reduce heterogeneity and produce polygons of danger levels in other weather

variables, such as rainfall, or for weather variables obtained from different numerical weather prediction models. Finally, we expect that since the proposed algorithm is based on purely image processing operations, it's performance will improve in phase with improvements in the weather forecasts.

Appendix A

Warning generation algorithm

The warning generation algorithm described in Section 1.3.2 can be implemented as

Algorithm 1 Generation of Warning Polygons

$w \leftarrow ones_like(p)$
for warning levels $i = 5, 4, 3, 2$ **do**
    $m = ones_like(w)$
    $m[p \geq i \text{ or } w > 1] \leftarrow p$
    $w_f \leftarrow f_f(m;\ t_f)$
    $w_f \leftarrow f_s(w_f;\ t_s)$
    $w \leftarrow max(w,\ w_f)$
end for
$w \leftarrow f_r(w;\ t_r)$

with the parameters defined in Table A.1.

1. The variable m is initialized with level 1 and the same shape as data preprocessing output.
2. The points of the current (i) or higher warning level occurring in p and w are saved in m. This information is necessary to generate the warning polygons for level i. When polygons of higher levels are saved in w, then these are included in iteration i to enable the generation of large enough polygons in level i.
3. Applying the morphological filtering function $f_f(t_f)$ to m forms contiguous polygons with reduced heterogeneity. As shown in Figure 2.3, first, a dilation is applied to connect singular points and form small areas. Second, an erosion reduces the heterogeneity (larger filter than first dilation to ensure singular enlarged points are removed). Third, dilating m to enlarge the before reduced area and make sure polygons of sensible sizes (as few as possible false alarms and misses, Cf. Table 2.2) are generated.
4. The median filtering $f_s(t_s)$ smooths out warning polygons of warning level i generated in the previous step. This reduces the heterogeneity from the last dilation because too small regions cancel out with large median filter sizes.
5. By saving the point-wise maximum of w and w_f, warning polygons of previous

iterations saved in w are kept, while adding warning polygons of the current warning level i (without overwriting saved warning polygons).

The for loop iterates over the warning levels from 5 to 2. For winter storms, a warning of level 5 or 4 is improbable to lie next to a level of 1. This is prevented with the selected iteration direction because in step two of the for loop, generated warning levels are stored in the temporary variable. This gives the chance in the next level $(i-1)$ to generate a neighboring polygon to the one of current level i. Iteration through level 1 is not necessary because w is initialized with ones.

All generated warning polygons of every iteration are saved in w. The boundary region around Switzerland (Cf. Equation 2.1) is cut off in variable w. This can produce small warning polygons at the boundaries of Switzerland. These are removed by applying the function $f_r(t_r)$ which changes areas with sizes below t_r to the surrounding warning level. This adaption is saved back in w where the output of the algorithm is saved in.

$f_f : \{1,2,3,4,5\}^{186\times131} \to \{1,2,3,4,5\}^{186\times131}$	morphological filtering
$f_s : \{1,2,3,4,5\}^{186\times131} \to \{1,2,3,4,5\}^{186\times131}$	median filtering
$f_r : \{1,2,3,4,5\}^{186\times131} \to \{1,2,3,4,5\}^{186\times131}$	remove too small polygons

$t_f \in \mathbb{N}^{4\times3}$	filter sizes for morphological operations (3 filter sizes per warning level (4))
$t_s \in \mathbb{N}^{4\times1}$	filter sizes for median filters (1 filter size per warning level (4))
$t_r \in \mathbb{N}$	threshold defining too small polygons
$w \in \{1,2,3,4,5\}^{186\times131}$	warning generated by algorithm

Table A.1: Explanation of functions and variables of Algorithm 1. Definition of p Cf. Equation 2.1. The functions f_f, f_s, and f_r map from one warning level per grid point ($\{1,2,3,4,5\}^{186\times131}$, where 186×131 is the dimension of the COSMO2-E grid used) to one warning level per grid point. The size corresponds to the shape of generated warning w. The filters' sizes and thresholds are natural numbers (specific numbers used see Table 2.1). The parameters' dimension is given by four iterations of the for loop (three filtering and one smoothing operations for each iteration).

Appendix B

Abbreviations

NWP	Numerical Weather Prediction
COSMO	Consortium for Small-scale Modeling
mMSL	Meters above mean sea level
POD	Probability of detection
FAR	False alarm ratio
SEDI	Symmetric extremal dependence index

Bibliography

Blake, Benjamin T., Jacob R. Carley, Trevor I. Alcott, Isidora Jankov, Matthew E. Pyle, Sarah E. Perfater, and Benjamin Albright (2018). "An Adaptive Approach for the Calculation of Ensemble Gridpoint Probabilities". In: *Weather and Forecasting* 33.4, pp. 1063–1080. DOI: 10.1175/WAF-D-18-0035.1.

Bolelli, Federico, Stefano Allegretti, Lorenzo Baraldi, and Costantino Grana (Oct. 2019). "Spaghetti Labeling: Directed Acyclic Graphs for Block-Based Connected Components Labeling". In: *IEEE Transactions on Image Processing* PP, pp. 1–1. DOI: 10.1109/TIP.2019.2946979.

Bradski, G. (2000). "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools*.

Ferro, Christopher A. T. and David B. Stephenson (2011). "Extremal Dependence Indices: Improved Verification Measures for Deterministic Forecasts of Rare Binary Events". In: *Weather and Forecasting* 26.5, pp. 699–713. DOI: 10.1175/WAF-D-10-05030.1.

Jain, Ramesh, Rangachar Kasturi, and Brian Schunck (Jan. 1995). *Machine Vision*. MIT Press and McGraw-Hill, Inc., Printed in USA, pp. 61–69, 122–124. ISBN: 978-0-07-032018-5.

Kain, John S., Steven J. Weiss, David R. Bright, Michael E. Baldwin, Jason J. Levit, Gregory W. Carbin, Craig S. Schwartz, Morris L. Weisman, Kelvin K. Droegemeier, Daniel B. Weber, and Kevin W. Thomas (2008). "Some Practical Considerations Regarding Horizontal Resolution in the First Generation of Operational Convection-Allowing NWP". In: *Weather and Forecasting* 23.5, pp. 931–952. DOI: 10.1175/WAF2007106.1.

Kropf, C. M., E. Schmid, G. Aznar-Siguan, S. Eberenz, T. Vogt, C. B. Steinmann, T. Röösli, S. Lüthi, I. J. Sauer, E. Mühlhofer, J. Hartman, B. P. Guillod, Z. Stalhandske, A. Ciullo, C. Fairless, P. M. (Mannie) Kam, wjan262, S. Meiler, R. Bungener, V. Bozzini, D. Stocker, and D. N. Bresch (Apr. 2022). *CLIMADA-project/climada_python:* version v3.1.2. DOI: 10.5281/zenodo.6475897.

McBride, John L. and Elizabeth E. Ebert (2000). "Verification of Quantitative Precipitation Forecasts from Operational Numerical Weather Prediction Models over Australia". In: *Weather and Forecasting* 15.1, pp. 103–121. DOI: 10.1175/1520-0434(2000)015<0103:VOQPFF>2.0.CO;2.

MeteoSchweiz (May 2021). *Beschreibung zu den Gefahrenstufen*. https://www.meteoschweiz.admin.ch/home/wetter/gefahren/erlaeuterungen-der-gefahrenstufen/wind.html.

MeteoSchweiz (2022). *COSMO forecasting system*. https://www.meteoswiss.admin.ch/home/measurement-and-forecasting-systems/warning-and-forecasting-systems/cosmo-forecasting-system.html. Accessed: 2022-01-01.

MeteoSchweiz, Warnkoordination (Sept. 2017). *Verfahren manuelle Warnverifikation*. Bundesamt für Meteorologie und Klimatologie MeteoSchweiz, Warnkoordination MeteoSchweiz, Urs Graf, MeteoSwiss, Meteorologist, private communication.

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

Planat (2021). *Winter storm National Platform for Natural Hazards*. https://www.planat.ch/en/knowledge-base/storm/winter-storm. Accessed: 2021-12-27.

Roberts, Brett, Israel L. Jirak, Adam J. Clark, Steven J. Weiss, and John S. Kain (2019). "PostProcessing and Visualization Techniques for Convection-Allowing Ensembles". In: *Bulletin of the American Meteorological Society* 100.7, pp. 1245–1258. DOI: 10.1175/BAMS-D-18-0041.1.

Schwartz, Craig S. and Ryan A. Sobash (2017). "Generating Probabilistic Forecasts from Convection-Allowing Ensembles Using Neighborhood Approaches: A Review and Recommendations". In: *Monthly Weather Review* 145.9, pp. 3397–3418. DOI: 10.1175/MWR-D-16-0400.1.

Toth, Zoltan, Eugenia Kalnay, Steven M. Tracton, Richard Wobus, and Joseph Irwin (1997). "A Synoptic Evaluation of the NCEP Ensemble". In: *Weather and Forecasting* 12.1, pp. 140–153. DOI: 10.1175/1520-0434(1997)012<0140:ASEOTN>2.0.CO;2.

Wilkinson, Michael and Jos Roerdink (Jan. 2009). *Mathematical Morphology and Its Application to Signal and Image Processing*. Vol. 5720. Springer-Verlag. ISBN: 978-3-642-03612-5. DOI: 10.1007/978-3-642-03613-2.

MeteoSchweiz
Operation Center 1
CH-8044 Zürich-Flughafen
T +41 58 460 99 99
www.meteoschweiz.ch

MeteoSvizzera
Via ai Monti 146
CH-6605 Locarno Monti
T +41 58 460 97 77
www.meteosvizzera.ch

MétéoSuisse
7bis, av. de la Paix
CH-1211 Genève 2
T +41 58 460 98 88
www.meteosuisse.ch

MétéoSuisse
Chemin de l'Aérologie
CH-1530 Payerne
T +41 58 460 94 44
www.meteosuisse.ch