

# Verification of radar-based hail detection algorithms with insurance loss data in Switzerland

MASTER'S THESIS

Faculty of Science

University of Bern

presented by

**Sandrine Morel**

2014

Supervisor:

Prof. Dr. Olivia Romppainen-Martius

Institute of Geography and Oeschger Centre for Climate Change Research

Advisor:

Luca Nisi

Institute of Geography, MeteoSwiss and Oeschger Centre for Climate Change  
Research



# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Motivation . . . . .	7
1.2	State of research . . . . .	7
1.3	Research questions and hypotheses . . . . .	9
<b>2</b>	<b>Data</b>	<b>10</b>
2.1	Radar data . . . . .	10
2.2	Insurance data . . . . .	12
<b>3</b>	<b>Method</b>	<b>14</b>
3.1	Data preparation . . . . .	14
3.2	Verification scores . . . . .	16
3.3	Setting lower thresholds on the number of claims . . . . .	19
3.4	Maximum and percentiles of POH and MESHS values . . . . .	20
3.5	Different adjustments of the datasets . . . . .	21
<b>4</b>	<b>Results</b>	<b>28</b>
4.1	Verification scores . . . . .	28
4.2	Frequency of claims per POH/MESHS values . . . . .	44
<b>5</b>	<b>Discussion</b>	<b>47</b>
5.1	Verification scores . . . . .	47
5.2	Discussion concerning the high FAR . . . . .	56
<b>6</b>	<b>Conclusion</b>	<b>58</b>
6.1	Outlook . . . . .	59
	<b>Appendices</b>	<b>61</b>
	<b>List of figures</b>	<b>71</b>
	<b>List of tables</b>	<b>75</b>



# Abstract

Severe hailstorms have a low probability of occurrence but are atmospheric phenomena that damage infrastructures, vehicles and crops considerably. In Switzerland, the impact is of high importance for the agricultural activities as well as for insurances. The present Master's Thesis focuses on the verification of two radar-based hail detection algorithms provided by MeteoSwiss; POH (Probability Of Hail) and MESHS (Maximum Expected Severe Hail Size). The study combines a 10-year radar-based hail climatology for Switzerland with motor insurance loss data for the time period 2003 to 2012. For the analysis, both insurance and radar datasets are analysed at the same spatial resolution (i.e. at a postal code area level). The verification of the two radar algorithms is mainly based on the Hit rate and on the False alarm ratio (FAR) calculated from a  $2 \times 2$  contingency table. In order to reduce the uncertainties due to the insurance and the radar datasets, adjustments are made, such as manual corrections of the insurance data or the application of a GIS-mask for traffic areas on the radar data. The insurance corrections strongly reduce the uncertainty due to the insurance data, while the mask does not reduce the uncertainty due to the location of the cars and does not improve the results. The results indicate that the FAR remains particularly high whatever adjustments are made and more uncertainties must be removed to make it reliable. The high Hit rate obtained with the POH algorithm indicates that hail events on the ground are generally well detected by the algorithms. However, data on the ground indicating the size of the hailstones would be necessary to accurately assess the quality of the MESHS algorithm.



# 1 Introduction

## 1.1 Motivation

Hail events are local atmospheric phenomena generated by strong thunderstorms that have a low probability of occurrence but lead to important impacts on infrastructure (e.g. solar panels, green houses, buildings), vehicles and crops. The related losses amount to several dozen million francs per year which makes hail one of the most costly natural hazard of central Europe [Berthet et al., 2012; Kunz et al., 2009; Kunz and Puskeiler, 2010; Kunz and Kugel, 2014; Mohr and Kunz, 2012]. Hail detection and forecasting systems need to be accurate in order to assess the vulnerability of the different objects and areas and to evaluate the cost of the events. Accurate radar-based hail detection and forecast could also allow for the launch of a hail suppression programme. The orography of Switzerland is an additional challenge as the strength and presence of meteorological events can strongly vary within a few hundred meters and as it also reduces radar and satellite visibility. Thus, the use of hail detection instruments and systems need to be adapted to this complex terrain. This topic becomes all the more important with climate change, as the changes might have an impact on the distribution and return period of hail events and on the size of the hailstones. This has already been shown by Kunz et al. [2009] who highlighted an increase of the number of hail days in the last 20 years in southwest Germany.

In his PhD study, Nisi [2013] compiled 12 years of meteorological data on hailstorms which occurred over the Swiss Alpine region, with the help of radar data. As the radars only provide an indirect estimation of the presence of hail in convective cells, it is necessary to have a ground-based observational dataset to validate the radar-based hail detection algorithms. However, there is no automatic hail detection system on the ground available in Switzerland. The present Master's Thesis will assess the quality and accuracy of two of the algorithms used by Nisi [2013], namely the POH (Probability Of Hail) [Foote et al., 2005; Waldvogel et al., 1979] and the MESHs (Maximum Expected Severe Hail Size) [Treloar, 1998]. This validation work will be based on insurance motor loss data, covering the entire territory of Switzerland for the years 2003 to 2012.

## 1.2 State of research

Several studies on hailstorms have already been carried out in order to assess the risk and damage of hail and to improve the identification of atmospheric conditions which might provoke hailstorms. Some of them are presented below. These studies are carried out for specific

areas only and are based on different radar-based hail detection algorithms. One of the major problems in hail data validation, whatever the data source, is the lack of ground-based observations. The latter are indeed very dependent on the population density of the different places.

The method mostly used to verify hail detection systems is based on insurance data [Betschart and Hering, 2012; Huntrieser et al., 1997; Kunz et al., 2009; Kunz and Puskeiler, 2010; Kunz and Kugel, 2014; Mohr and Kunz, 2012; Schuster et al., 2006; Skripniková, 2013]. The damage observed on the ground can be compared to the hail detection by the radar algorithms and confirm or invalidate the detection. However, many uncertainties must be taken into consideration, such as the different spatial resolution between the hail detection and observation, the disparities in the population coverage over the area of study or the errors that can be present in the insurance dataset. Other reasons will be discussed more in detail in further chapters.

It is also possible to use hail reports [Delobbe et al., 2005; Betschart and Hering, 2012]. These are manually collected during hail events and provide information on the size of the hailstones as well. However, as there is no way to automatically compile such reports, it is time consuming and quality can vary greatly between different areas. Data collected through public media can be useful if there is a way to detect and collect it automatically.

Another possibility to verify hail detection systems is to use hailpads [Berthet et al., 2012]. The latter can be more regularly spread on the territory and reduce the uncertainty due to the location of the population. However, this method is expensive and is still not available in many countries, including Switzerland.

All these methods help to verify hail detection systems that are usually used for hail forecasting. These hail detection systems are based either on the assessment of convective indices and parameters or on radar-based hail detection algorithms.

Regarding hail forecast at a local level, it is possible to investigate the meteorological situations that often lead to hail events [Kunz et al., 2009; Merino et al., 2013; Berthet et al., 2012]. This could allow to forecast a possible hail event when the situation is forming and to proceed to hail suppression. However, the best atmospheric parameters to take into account to detect and forecast hail are still not well known. It is thus interesting and necessary to assess the accuracy of the different methods to know which one fits which orography and atmospheric conditions best.

Holleman et al. [2000] found that the method used by Waldvogel et al. [1979] (the POH algorithm) was the best compared to the four other ones assessed in a study they carried out during 15 days of severe weather in the Netherlands. They used eye-witness reports and damage reports from the agricultural insurance networks for their ground-based data. In another study carried out by Kunz and Kugel [2014], five different hail detection algorithms were evaluated with the help of data from a building insurance company. This study covered southwestern Germany for the years 1997-2012. They found out that for large parts of the investigation area, the Waldvogel algorithm was able to well reproduce hail signatures. It produces a high Hit rate but also create quite a few false alarms. The two algorithms performing best for the identification of hail signatures are the Hail Detection Algorithm HDA [Smart and Alberty, 1985] and the Probability of Severe Hail POSH [Witt et al., 1998].

While Betschart and Hering [2012] validated the data obtained from the radar network



of MeteoSwiss (POH, MESHS) with the help of insurance records for the years 2009 and 2010, and with observations collected on the Internet for the year 2011, the present Master's Thesis will only make use of the insurance records but in a more comprehensive way. The main advantage of insurance loss data is that it provides a better coverage than reported observations which can be collected through the Internet.

### 1.3 Research questions and hypotheses

Only few studies concerning radar-based hail detection algorithms have been conducted for Switzerland, and none of them covered the entire territory for a long time period. The present Master's Thesis is thus carried out in order to close this gap by using datasets available over the whole territory of Switzerland and for a time period of 10 years. As mentioned by Kunz and Puskeiler [2010], insurance data provides a good overview of the damaging hailstorms, but varies according to the population density of the regions. The radar data, on the other hand, allows to detect hailstorm tracks, but does not allow to know if hail indeed occurred on the ground. Both datasets contain uncertainties, but used together they promise interesting and reliable results.

The main objective of the present study is the evaluation of the two radar-based hail detection algorithms POH and MESHS from the MeteoSwiss radar network with the help of insurance loss data based on cars.

The following questions are addressed:

- How well do the radar-based hail detection algorithms predict the probability of hail on the ground ?
- Above which POH and MESHS values does important damage occur ?

Two hypotheses are formulated to give guidelines to the study:

- The POH algorithm takes into account a wide range of hail sizes, sometimes even considering hailstones lower than 2 cm and graupel. It is expected that the POH algorithm gives better results than the MESHS algorithm when hail indeed occurred. However, it is also expected that the POH algorithm produces more false alarm cases than the MESHS algorithm.
- Damage on cars is expected to increase with increasing POH values.

In the following chapters, the data will first be presented (Chapter 2) and followed by a description of the method in Chapter 3. The results will then be described (Chapter 4) and discussed (Chapter 5). The last chapter (6) will draw a conclusion and present an outlook on the further work to be carried out.

## 2 Data

The present study uses two datasets: hail information retrieved by the two radar-based hail detection algorithms (POH and MESHS) and the insurance loss data from the Mobiliar insurance. Both datasets contain uncertainties, but used together they bring nevertheless interesting and reliable results. The extended convective season between the 1<sup>st</sup> of April and the 30<sup>th</sup> of September 2003-2012 is considered. This period has been selected as most of the severe thunderstorms occur at this time of the year in this region and it is considered as the hail “season”.

### 2.1 Radar data

#### 2.1.1 MeteoSwiss radar network

The MeteoSwiss radar network consists of three radars located in la Dôle (Jura of the Vaud), on the Albis (Zurich) and on the Monte Lema (Ticino). The radars are volumetric C-band Doppler radars operational since respectively 1995, 1994, 1993. The antenna scans every 5 minutes a full circle at 20 elevation angles ranging between  $-0.3^\circ$  and  $40^\circ$  up to a maximum range of 230 km [Germann et al., 2006]. All three radar stations have been replaced between 2011 and 2012 with state-of-the-art dual-polarized Doppler radar stations to insure a good quality of the data. The spatial resolution of the radar-based hail detection algorithms is  $1 \times 1 \text{ km}^2$  and the temporal resolution is 5 minutes. From this data, aggregations per hour or per day are possible [Germann et al., 2006; Hering et al., 2004; Joss et al., 1997].

In the present study, daily maximum records are used and retrieve the maximum value of the radar-based hail detection algorithms for every pixel. Because of the renewal of the radar network, the availability of the radars was limited between 2011 and 2012. Radar information is thus partly missing for this time period. This should not strongly affect the present analysis as the radars were offline only for a few days [Luca Nisi, personal communication, June 2014].

#### 2.1.2 Radar-based hail detection algorithms

Hail detection at MeteoSwiss is done using two different algorithms which provide information on the probability of hail and on the hailstone size: Probability Of Hail (POH) and Maximum Expected Severe Hail Size (MESHS). The POH algorithm is operationally available since 2008 and the MESHS algorithm since 2009. The present study is thus based on data reprocessing of these two algorithms for the period 2002-2013 [Nisi, 2013]. Both algorithms provide a

graphical output every 5 minutes and can also be presented as daily composite hail detection maps (Fig. 2.1). This daily composite is based on the integration of the maximum detected hail signal for each day and each pixel from 00.00 to 23.55 UTC [Betschart and Hering, 2012].

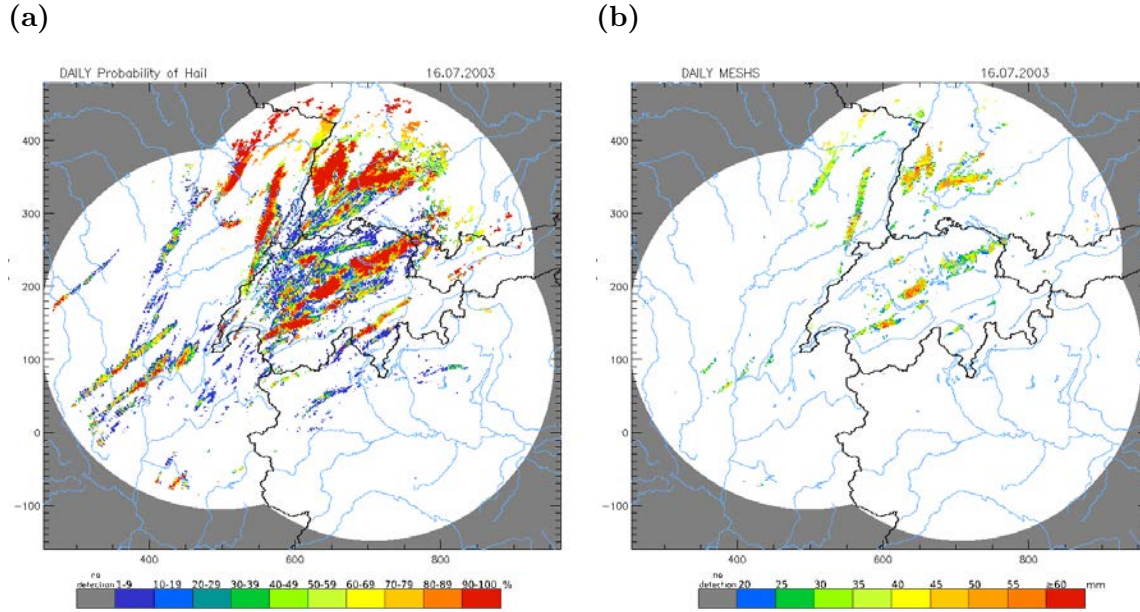


Figure 2.1: Daily maps for the POH (a) and MESHS (b) algorithms [MeteoSwiss].

The expected probability of hail on the ground is a non-linear relation based on the difference of altitude  $\Delta H_{45} = H_{45} - H_0$ ,  $H_{45}$  being the altitude of the 45 dBZ (decibels relative to Z, Z being radar reflectivity factors) and  $H_0$  the altitude of the freezing level (Fig. 2.2). Hail is expected on the ground when  $\Delta H_{45} > 1.4$  km.  $\Delta H_{45} > 6$  km means 100 % probability of hail occurrence [Waldvogel et al., 1979; Foote et al., 2005].

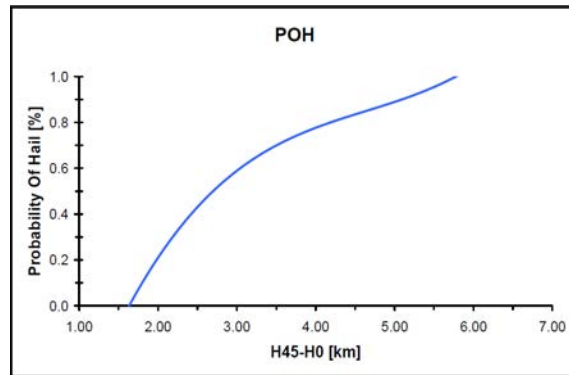
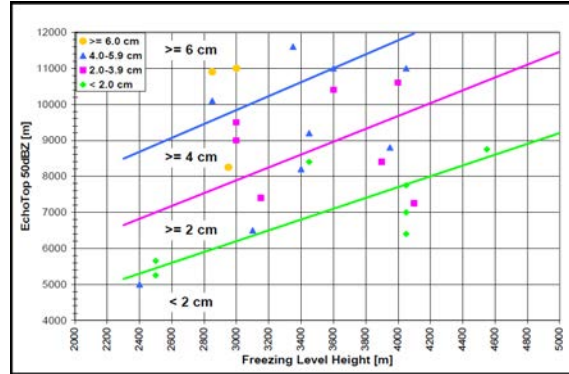


Figure 2.2: POH on the ground according to the difference of altitude between the height of the 45 dBZ and the freezing level [Waldvogel et al., 1979; Foote et al., 2005].

The MESHS algorithm provides information about the maximum hailstone diameter that can be expected on the ground. It is determined using an empirical relation based on the difference of altitude  $\Delta H_{50} = H_{50} - H_0$ , with  $H_{50}$  being the altitude of the 50 dBZ and  $H_0$  the

altitude of the freezing level, and is illustrated by Fig.2.3 [Treloar, 1998]. The MESHS values are retrieved in 0.5 cm classes and the expected sizes smaller than 2 cm are not computed. In this study, the MESHS algorithm is used with classes of 10 mm as the uncertainty relative to the expected size is about 1 cm [Luca Nisi, personal communication, April 2014].



**Figure 2.3:** Nomogram of MESHS on the ground according to the difference of altitude between the height of the 50 dBZ and the freezing level [Treloar, 1998].

For both algorithms,  $H_0$  is extracted from the COSMO-CH analysis (<http://cosmo-model.org/>), which is used for atmospheric prediction. COSMO-CH is a non-hydrostatic, regional, high-resolution model operated and further developed by MeteoSwiss [Luca Nisi, personal communication, July 2014]. In this study, the POH and MESHS maximum values per day are used (Fig. 2.1a and 2.1b).

## 2.2 Insurance data

The insurance dataset used in the present study consists of information on hail damage done to motorcars. It has been provided by the Mobiliar insurance in Switzerland for the years 2003 to 2012. The information on total losses is available at a daily resolution. The spatial resolution consists of the postal code areas. The claims are listed in the following way: a form to report the damage to the insurance has to be filled out when a driver observes a damage on his car. In this form, some information should be mentioned, such as the date of the damage, the postal code area where the damage occurred and the postal code area of the place of residence of the customer.

The advantage of such a dataset is that it provides information for almost all the populated places as about 80 % of the households in Switzerland have at least one car. Even if each of the cars are not insured by the same insurance company, the Mobiliar insurance coverage is still representative of the Swiss population. Moreover, cars are more sensitive to small hailstones than buildings, there is therefore more information related to car damage.

### 2.2.1 Limitations of the insurance dataset

This dataset has limitations and uncertainties which must be considered and which will be described into detail in Chapter 3:

- The cars are moving objects and it is thus sometimes difficult to know precisely where the damage occurred.
- The dataset is dependent on the population distribution, as more cars are expected to be present in very populated places than where nobody lives. The number of cars insured by the Mobiliar can also vary a lot between postal code areas.
- Errors can also come from the forms filled out by the customers. A wrong date or a wrong postal code area can be indicated.
- Finally, a limitation is due to some uncertainties in the postal code areas during the years 2003-2012.

# 3 Method

The different steps of the method shown on Fig. 3.1 will be presented in the following sections.

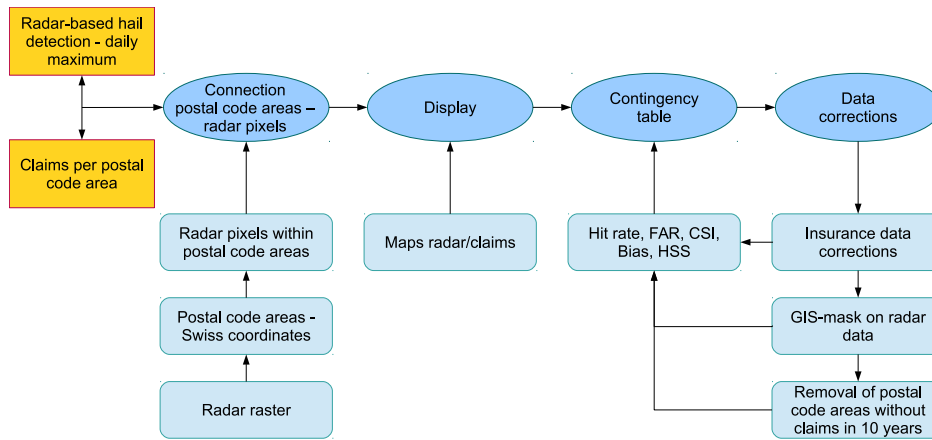


Figure 3.1: Flow chart of the methodology.

## 3.1 Data preparation

### 3.1.1 Uncertainties due to the postal code areas in the insurance dataset

The postal codes given in the insurance reports to indicate where the damage occurred did not always correspond to a precise locality. The data needed to be updated to remove or correct the incorrect postal codes. Several cases where the postal codes were modified or removed are presented below.

First, some insurance reports indicated a post office box as location of the hail event that damaged the car. However, the post office boxes do not correspond to a specific locality but to a post office. It is assumed that when the code of a post office box is indicated, it probably corresponds to the customer’s address or to the locality of the post office. It is therefore rarely the place where the hail damage actually occurred. The customer probably wrote this code because he did not know the postal code of the place where he was when the hail damage occurred. As the uncertainty is big and as the indication of a post office box was only present in 200-300 of over 65,000 reports, it has been decided to remove these claims from the dataset used in this study.

Second, some of the reports indicated the postal code of a company, but this kind of postal code is not officially recognized within the standard postal codes. Thus, when the company did not have any branches, the postal code of the company location was taken. This was possible assuming that the customer was probably working there and did not know the official postal code of the place but only the one of the company. If the company did have branches in other postal code areas, the report was simply removed from the dataset, as this was the case for a few claims only.

Another uncertainty arose with postal codes corresponding to a general region but not to a precise locality within this region. Some reports contained these postal codes, probably because the customer did not precisely know in which specific part of the region he was. In this case, the claims were distributed among the existing postal codes of the region. The postal codes assigned in priority were the ones in the city centers, followed by the ones where the main roads were located, as this is where it is expected to find most of the cars. However, this method is based on an assumption as the cars could have been somewhere else in the area. About 4000 claims were filled in with a general postal code, and that is why it was decided to find a solution to still keep them, even if an uncertainty was added to the dataset.

It also happened that some of the postal codes mentioned in the insurance reports did not exist anymore because they were grouped together with other postal codes or because the code just changed. When it was possible, the postal codes were updated with the new postal code. The postal codes that changed and were not updated were kept if damage occurred before the year of change. After the year of change, they do not indicate anymore damage as they are all counted in the new postal code. Only very few of them were not classifiable anymore and were removed from the dataset.

Finally, an uncertainty also occurred with borders of the postal code areas, as they might have changed during the last 10 years, but this is thought to cause a very small uncertainty and it is therefore dismissed. This study is based on the assumption that no changes occurred in the borders as the investigation would require a huge amount of work and should not strongly change the results. The postal codes that did not appear in the four previous categories were assumed to be correct.

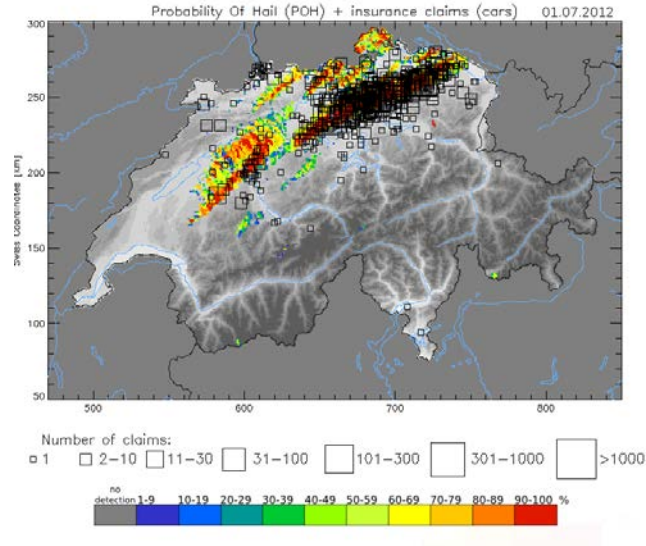
### 3.1.2 Linking the postal code areas with the radar pixels coordinates

The first step in the data preparation is to link both the radar and insurance datasets together at the same spatial resolution to allow for analysis and comparison (Fig. 3.1). The radar data provides one value for each pixel (swiss coordinates) of  $1 \times 1 \text{ km}^2$ , while the insurance data provides the number of claims for each postal code area. That is why the method used allows to know which pixels and their corresponding radar detection values are within which postal code area.

However, it often happens that parts of several different postal code areas are together in one pixel. In these cases, the pixel is considered to belong to all the postal code areas present in it, without taking into account the size of the area of each postal code area within the pixel.

Fig. 3.2 shows an example of visualisation of the radar and insurance datasets brought together. The claims (black squares) are given per postal code area and the radar values are

given per pixel. It gives a first idea of the correspondence between the two datasets.



**Figure 3.2:** POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 1<sup>st</sup> of July 2012 [graphics: L. Nisi].

## 3.2 Verification scores

The method used to assess the quality of the radar-based hail detection algorithms is basically the same as for a forecast verification. It is based on the correspondence between measurements in altitude and observations on the ground. In order to facilitate the analysis, the different POH and MESHS values are converted into binary data. Using thresholds according to the POH/MESHS values, it is then possible to discriminate between hail detection and no hail detection. For example, with a POH threshold of 60 %, all the values between 0 and 59 % are considered as “no radar detection”, while the values between 60 and 100 % are considered as “radar detection”.

The common method in meteorology to verify the forecast or, in the present case, the radar detection is to use a contingency table and its related ratios and skill scores [Brimelow and Reuter, 2006; Huntrieser et al., 1997; Kunz and Kugel, 2014; Skripniková, 2013; Wilks, 2006].

### 3.2.1 The contingency table

In the following chapters, the term “detection” is used to discriminate between a detection and no detection of hail from the radar. If the POH value exceeds the POH threshold previously set, it is a positive detection. If the POH value is below the threshold, it is a negative detection. The term “observation” is used to discriminate between an observation or no observation of hail on the ground, corresponding to claims in the present study. If the number of claims per postal code area exceeds the threshold of claims previously set, it is a positive observation. If



the number of claims per postal code is lower than the threshold of claims previously set, it is a negative observation.

The contingency table is based on the joint distribution of the detections and observations [Wilks, 2006]. In the present case, a  $2 \times 2$  contingency table is used (Fig. 3.3). The entries are the positive/negative detections and the positive/negative observations.

Detected \ Observed	Yes	No	
Yes	a	b	a+b
No	c	d	c+d
	a+c	b+d	n=a+b+c+d

**Figure 3.3:**  $2 \times 2$  contingency table for the verification of radar-based hail detection algorithms, based on the positive/negative detections and positive/negative observations.

As shown on Fig. 3.3, the  $2 \times 2$  contingency table shows the relationship, or correspondence, between the counts of detected/observed pairs (letters  $a-d$ ) and their marginal totals, in absolute terms. The letter  $a$ , or *hit*, indicates the cases where there is a radar-based hail detection and an observation on the ground. The letter  $b$ , or *false alarm*, indicates the cases where there is a radar-based hail detection but no observation on the ground. The letter  $c$ , or *missed*, indicates the cases where there is no radar-based hail detection but an observation on the ground. And the letter  $d$ , or *non-event*, indicates the cases where there is no radar-based hail detection and no observation on the ground. The sample size  $n$  is equal to all the cases added together. The marginal totals show the total number of times of positive ( $a + b$ ) or negative ( $c + d$ ) detections, without taking into account the fact that an event was indeed observed or not, and conversely for the positive ( $a + c$ ) and negative ( $b + d$ ) observations [Wilks, 2006].

The contingency table is necessary to calculate verification scores. These will help to assess the accuracy of the radar-based hail detection algorithms. The verification scores are calculated for all Switzerland as well as for every postal code area.

As presented by Wilks [2006] and used in several studies [Betschart and Hering, 2012; Brimelow and Reuter, 2006; Huntrieser et al., 1997; Kunz and Kugel, 2014; Skripniková, 2013], the two main ratios used to assess the accuracy of the radar algorithms are the Hit rate and the False Alarm Ratio (FAR). They vary between 0 and 1. For a good detection, the Hit rate should be as high as possible while the FAR should be as low as possible. The three other scores usually used to evaluate the radar detection are the Critical Success Index (CSI) [Betschart and Hering, 2012; Brimelow and Reuter, 2006; Huntrieser et al., 1997; Kunz and Kugel, 2014; Skripniková, 2013], the Bias [Betschart and Hering, 2012] and the Heidke Skill Score (HSS) [Brimelow and Reuter, 2006; Kunz and Kugel, 2014]. Their results give an idea of the general accuracy of the radar-based detection algorithms. All these results are scalar measures and therefore provide clear information on the accuracy of the radar detection.

### Hit rate

Adapted to the present study, the Hit rate corresponds to the ratio of correct radar detections (*hit*) divided by the total number of times a hail event has indeed been observed on the ground

(positive observations). It can also be seen as the fraction of cases when an event has been observed and detected on all the observed cases. The Hit rate is defined as

$$H = \frac{a}{a + c}. \quad (3.1)$$

A high Hit rate means that there is an accurate detection by the radar of the events actually occurring on the ground. A low Hit rate, on the contrary, means that many claims on the ground are not captured by the radar.

## FAR

The FAR is a conditional relative frequency of the positive detections that can be shown as

$$FAR = \frac{b}{a + b}. \quad (3.2)$$

It corresponds to the fraction of positive detections that turn out to be wrong, or at least without any ground truth (*false alarms*). High values of FAR mean that the radar often makes a detection without any corresponding observation on the ground. On the contrary, low values of FAR mean that most of the radar detection cases are consistent with observations on the ground. It is therefore necessary to have a dataset of observations on the ground as large as possible to cover a maximum of the studied area. As already observed in several studies [Betschart and Hering, 2012; Kunz and Kugel, 2014], a lack of ground truth could distort the results, as a high FAR could either mean that the radar often detected hail wrongly or that it detected hail accurately but nobody was there to see whether it actually occurred or not.

## CSI

A simple index used to characterise the accuracy of the radar detection is the CSI:

$$CSI = \frac{a}{a + b + c}. \quad (3.3)$$

This is a useful index when the event to be forecast occurs substantially less frequently than the nonoccurrence [Wilks, 2006]. Therefore, it makes sense to use it in the case of hail detection, where events are often very localised. The CSI corresponds to the number of correct positive detections (*hit*) divided by the total number of cases when an event was detected and/or observed (*hit, false alarms, missed*). The CSI ranges between 0 and 1, 0 being the worst score and 1 the best score.

## Bias

According to Wilks [2006], the Bias (B) is defined as

$$B = \frac{a + b}{a + c} \quad (3.4)$$

and measures the correspondence between the average detection and the average observation. It corresponds to the ratio of the positive detections to the positive observations. A Bias of

1 means that the detection is unbiased and that the event was detected the same number of times that it was observed. However, the result does not give any information on the correspondence of particular cases, this is only an overall result based on all the cases included in the dataset. A Bias greater than 1 indicates more detection cases than observations and is called “overdetection”. Conversely, a Bias smaller than 1 indicates less detection cases than observations and is called “underdetection”.

## HSS

The HSS corresponds to the Proportion Correct (PC) that would be achieved by random detections that are statistically independent of the observations [Wilks, 2006]. The PC, on which relies the HSS, corresponds to

$$PC = \frac{a + d}{n}, \quad (3.5)$$

while the HSS has the following form:

$$HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)}. \quad (3.6)$$

Therefore, a perfect detection would have  $HSS = 1$ , a detection equivalent to the reference detection would have  $HSS = 0$  and a detection worse than the reference detection would have a negative score [Wilks, 2006].

## 3.3 Setting lower thresholds on the number of claims

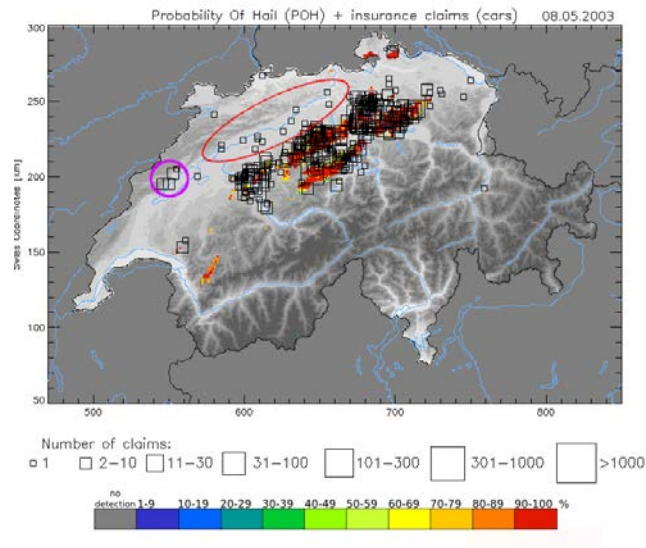
As described in Chapter 2.2.1, the insurance data contains some uncertainties. In order to remove part of the uncertainties due to incorrect dates or locations, a threshold for the claims is applied. Generally, a hail observation on the ground is defined as the presence of at least one claim in a postal code area. With a threshold applied on claims, a hail observation on the ground is then defined as the presence of a minimal amount of claims, equal to the threshold, in a postal code area. For example, with a threshold of 2 claims, only the postal code areas with at least 2 claims in a particular day are considered to contain hail observations on the ground (positive observations). If only one claim is present in the postal code area, it is considered as an error and not as a hail observation on the ground (negative observations).

The verification scores presented above (Chapter 3.2.1) are expected to vary according to the threshold applied on claims. A relevant threshold should allow to increase the Hit rate by removing wrong claims. A threshold too high might reject too much valid data (*hit*) and increase the FAR, while a threshold too low might keep too much invalid data (*missed*) and reduce the Hit rate.

Faced with similar errors in their datasets, Kunz and Puskeiler [2010] used in their study a lower threshold of 50 claims and 100'000 euros per day over the entire area of study (southwest Germany). For an analysis at a smaller scale, a lower threshold of 4 claims per day over a single postal code area has been used. In a further study over southwest Germany, Kunz and Kugel [2014] used a threshold of 10 claims per day. Skripniková [2013] defined a pixel as

“hail pixel” when more than 3 claims and at least 100 euros of hail damage were found in the corresponding postal code area for one day. In the present study, a lower threshold by postal code area only is used. Three thresholds are set up to compare their respective results and determine which one gives the best results without losing too much information that could be correct. The selected thresholds are 1, 2 and 5 claims per postal code area. A threshold of 1 claim takes all the claims into account, while the thresholds of 2 and 5 claims consider that, if few claims are observed in each single postal code area, these ones are in reality incorrect data.

Only one claim per postal code area per day is assumed to be an erroneous value, given the low probability of only one car being damaged by hail. And too much information is expected to be lost with a threshold of 5 claims. The threshold of 2 claims is expected to give the best and most accurate results with the lowest uncertainty. This is however a rough assumption as it depends on the size of the postal code area, on the number of insured cars and on the area hit by the hailstorm. As shown on Fig. 3.4, several cases with only one claim in the postal code area are present and often without any radar detection (red circle). The use of a threshold of 2 claims would allow to remove these potentially incorrect cases. There are also several postal code areas having between 2 and 10 claims but sometimes without any radar detection (purple circle). By using a threshold of 5 claims, the postal code areas having only between 1 and 4 claims would not be considered as hail observations on the ground. It is expected that these cases mostly take place where there is no radar detection, while the ones with more than 4 claims take place where there is a radar detection.



**Figure 3.4:** POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 8<sup>th</sup> of May 2003 [graphics: L. Nisi].

### 3.4 Maximum and percentiles of POH and MESHS values

Using a spatial resolution by postal code area generates an uncertainty according to the distribution of the POH and MESHS algorithms values. The size of the postal code areas is variable

and it can happen that only few pixels in the postal code area have a high POH/MESHS value while the other ones have much lower values. This might increase the FAR as claims are expected, even if only few pixels have a high POH/MESHS value. Moreover, these pixels might be over non-populated places. The present study mainly considers maximum radar detection algorithms values per postal code area as these are the ones expected to be related to most of the damage. The value per postal code area used in the analysis will thus be the highest one, even if most of the postal code area has low POH/MESHS values. To weight the high values, percentiles are also used and should allow to reduce the FAR. Indeed, if only few pixels have high POH/MESHS values within the postal code area, these pixels will not or less be taken into account with the use of percentiles. The 50<sup>th</sup> (median), the 75<sup>th</sup> and the 85<sup>th</sup> percentiles have been selected as they do not take into account the extreme high values of the postal code areas.

## 3.5 Different adjustments of the datasets

### 3.5.1 Insurance data corrections

The insurance dataset contains two major uncertainties: the first one is linked to the location of the hail event and the second one is linked to the date. The postal code usually corresponds to the location of the hail event, and the date of the damage on the car in the insurance report usually corresponds to the date of the hail event. However, some dates or postal codes might have been inaccurately indicated which brings errors to the dataset. As said in Chapter 3.3, the same kind of error was observed in several other studies but the method used to remove these inaccuracies was limited to the application of thresholds of claims and on losses amount [Kunz and Puskeiler, 2010; Kunz and Kugel, 2014; Skripniková, 2013]. However, in the present study, some cases with more than one claim in one postal code area are observed without any radar detection in the surrounding area. The correction of the erroneous locations has already been done during the preparation of the data (Chapter 3.1.1). That is why it has been decided to proceed with a manual correction of the erroneous dates only. The detection and correction of such cases are done manually and subjectively using the POH dataset.

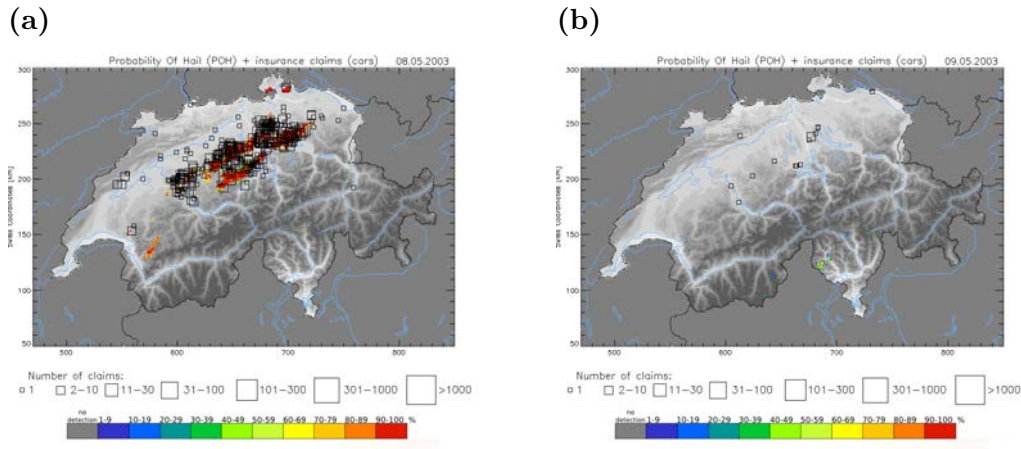
The following criteria are used for the correction:

- i) If on a day no POH values are observed where claims are located but POH values are observed on the previous or next days at that place, the claims are moved to the day of the hail event. The claims that are moved to that day are generally taken from the first or the second day before and after the event, but this limit is extended to five days before and after if several claims are on the same trajectory as the hail event (Fig. 3.5).
- ii) The claims are moved to the closest day with a hail event at that place or, if there are two close days, to the one with the highest POH values.
- iii) If on a day no POH values are observed where claims are located but POH values are observed one month earlier or later at that place, the claims are moved to the month of the hail event. The lag period considered between claims and the correct day of the month is one month. This is the case for the claims on the 25<sup>th</sup> of June 2006 without

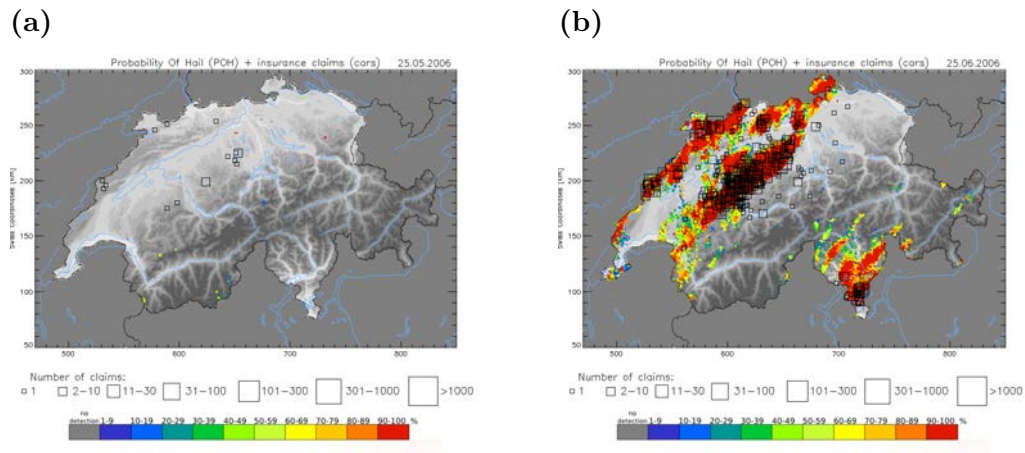
any POH values but following exactly the same pattern as the hail event on the 26<sup>th</sup> of July 2006, one month later (Fig. 3.6). Such cases are probably due to digitizing errors.

- iv) Not all the claims of one day are moved if some are not corresponding to any POH values in the surrounding days.
- v) If individual and remote claims are present without any POH values in this area in the surrounding days, they are removed (Fig. 3.7).
- vi) If several claims are present at a same location without any POH values in this area in the surrounding days, the precipitation radar detection is used to verify if there was at least precipitation over the area of the damage (Fig. 3.8). Hail may occur when there is a strong rainfall. Therefore, when a strong rainfall is detected in the area of the claims, the latter are kept in the dataset and removed when there was no or low radar detection, in order to reduce the bias due to wrong or inaccurate insurance data. There are about ten cases for which the verification with precipitation radar detection was done and about half of these were not corroborated with the radar data.
- vii) Many claims are listed on the first day of a month without any POH values (Fig. 3.9). This is probably due to the fact that reports with only the month as date were added to the first of the month to have a valid date. In these cases, as it is not possible to know to which day the claims belong, they are removed if there are no POH values on that day at the location of the claims.

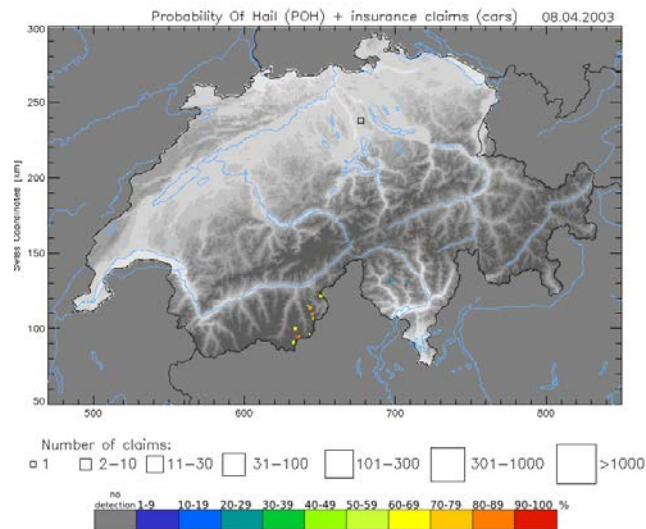
This process of verification is very subjective as it is done manually but it allows for the detection of mistakes and patterns in the different datasets. It also allows to have a more accurate insurance dataset than without any correction. The present dataset, on which corrections have been made, is named “dataset with insurance corrections” in the rest of this study.



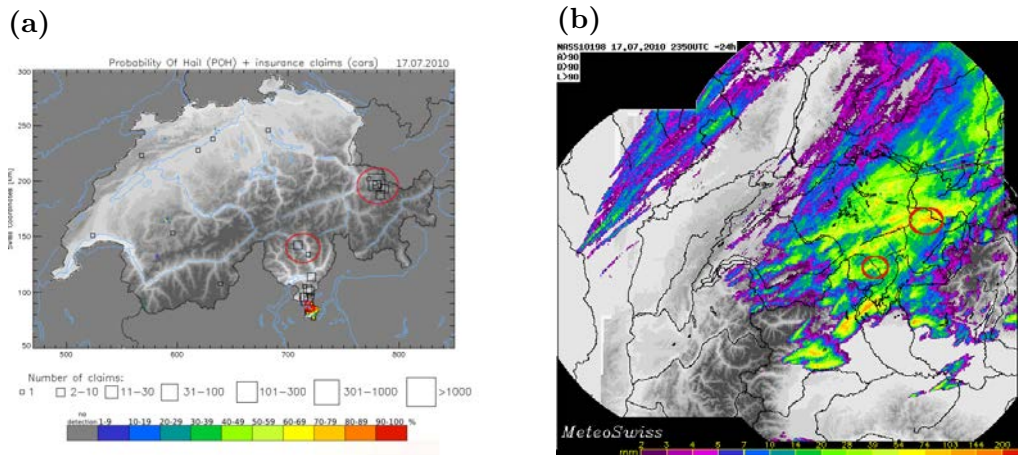
**Figure 3.5:** POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 8<sup>th</sup> of May 2003 (a) and on the 9<sup>th</sup> of May 2003 (b) [graphics: L. Nisi].



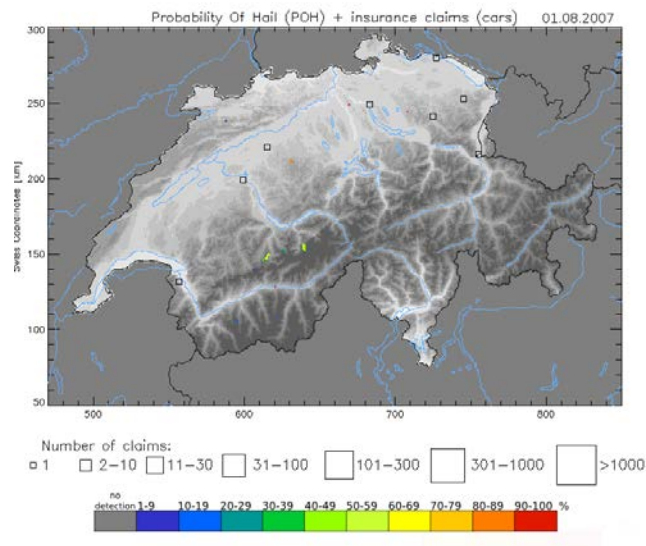
**Figure 3.6:** POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 25<sup>th</sup> of May 2006 (a) and on the 25<sup>th</sup> of June 2006 (b) [graphics: L. Nisi].



**Figure 3.7:** POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 8<sup>th</sup> of April 2003 [graphics: L. Nisi].



**Figure 3.8:** POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 17<sup>th</sup> of July 2010 (a) [graphics: L. Nisi] compared to the precipitation radar detection using daily maximum values (b) [MeteoSwiss].



**Figure 3.9:** POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 1<sup>st</sup> of August 2007 [graphics: L. Nisi].



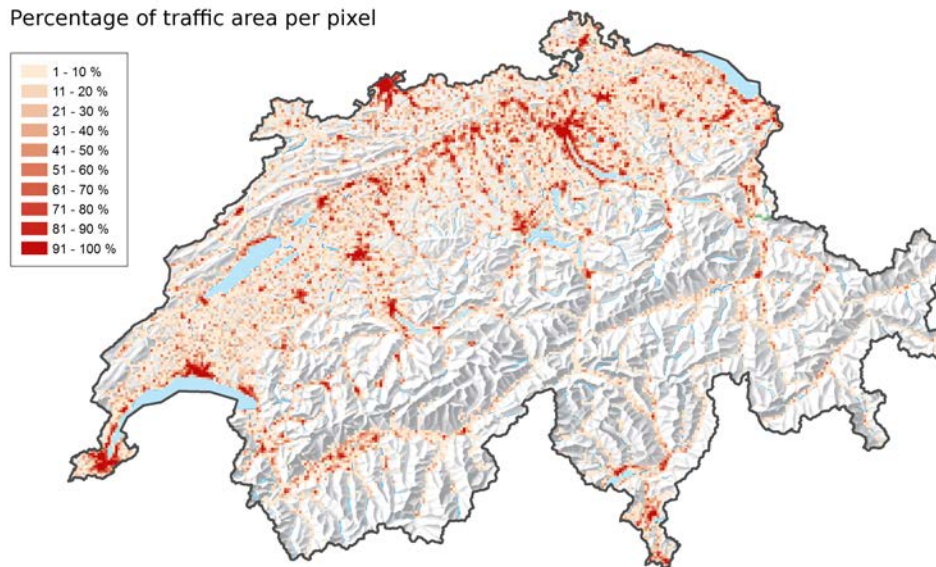
### 3.5.2 Adjustment with the GIS-mask on traffic areas

Some areas in Switzerland have a very low population density where almost no roads are present. This could explain the lack of insurance claims on the ground, as very few people are expected to live in these areas and therefore few cars should be present.

To remove most of the areas where cars are not expected to be, such as forests or mountains, a mask is applied on the dataset with insurance corrections and concerns only the radar data. This means that some pixels will lose their radar value but all the claims will stay where they are. The roads kept in the mask are those with high traffic. Moreover, including small roads would not be consistent with the spatial resolution of the radar. The other items correspond to areas where cars are expected to be frequently found. The following items are kept in the mask as areas where cars are expected:

- Motorways and their approach ramps;
- Motorways with only one direction for the circulation and their approach ramps;
- Semi-motorways and their approach ramps;
- Road of 10.20 m wide;
- Roads between 6 and 10 m wide and their approach ramps;
- Roads between 4 and 6 m wide and their approach ramps;
- Motorway and semi-motorway junctions;
- End of motorways and semi-motorways;
- Motorway and semi-motorway interchanges;
- Crossroads with approach ramps;
- Loading of cars on a railway;
- Road passes;
- Rest areas;
- Fast-food and filling stations;
- Tollbooths;
- Urban areas;
- City-centers;
- Road connections.

The mask is used in a format taking into account the percentage of the presence of the mask items in each pixel. The pixels with a percentage of mask items lower than the threshold set are considered as pixels where cars are not expected and their radar value is set to zero. For example, a mask using 5 % threshold considers all the pixels with at least 5 % of the area occupied by an item of the mask. When the radar detection is equal to zero, this is the same as if no hail occurred at this place and thus no damage is expected to be declared. Fig. 3.10 shows the pixels of the mask over Switzerland according to the percentage of traffic area. In the present study, the percentages of 1, 5 and 50 % are used for the analysis. The mask 1 % takes all the pixels in the mask into account. The masks 1 and 5 % are expected to give results close to each other. The mask 50 % strongly reduces the traffic areas and is expected to give results quite different from the other masks and from the datasets without the application of the mask. The mask 50 % was selected to see the importance of small roads and if the related uncertainty was important.



**Figure 3.10:** Mask applied on the dataset and showing the percentage of traffic area per pixel for Switzerland [graphics: L. Thomi].

The datasets that are created with the the application of the masks are named “datasets with the masks” in the rest of this study. The respective names of the masks are “mask 1 %”, “mask 5 %” and “mask 50 %”.

### Errors in the masks

When using the mask, it is important to take into account some errors and uncertainties present in it. First, two different datasets (swissTLM3D and VECTOR200) from Swisstopo were used together to create the mask, although they do not have the same spatial resolution. This adds an uncertainty to the mask as similar items were used from both datasets and did not juxtaposed. Some roads were thus sometimes counted twice. Moreover, artefacts were present in one of the dataset, adding some pixels to the mask even if they did not contain

any of the items. These two errors must be taken into account in the results. They can have an influence, mainly on the FAR.

### 3.5.3 Removing postal code areas without claims in a 10-year period

An other adjustment of the dataset is the removal of 819 postal code areas where no claims have been registered in 10 years. This adjustment is based on the dataset with insurance corrections. Several reasons could explain this absence of claims, such as the lack of insured cars in these postal code areas, the possible small size of the postal code areas, the very small amount of cars in these regions or the fact that no hail with hailstones larger than 2 cm fell there. A deeper investigation of these reasons would be very interesting but is out of the scope of the present study. This removal is expected to reduce the FAR without affecting the Hit rate, as neither claims nor radar detection where claims occurred were removed. For comparison, this adjustment is also applied on the dataset using the mask 1 %.

In the rest of this study, the datasets in which postal codes have been removed are called “datasets *removed postal codes without claims*”.

# 4 Results

## 4.1 Verification scores

### 4.1.1 Original dataset

The original dataset, without any corrections or adjustments, contains 1053 hail days for the time period 2003-2012 with a total of 66'536 claims.

#### POH

The Hit rate decreases as function of increasing POH thresholds (Fig. 4.1). The decrease is strongest between the POH thresholds of 80 % and 100 %. Using a threshold of 2 or 5 claims (red and yellow lines), the decrease of the Hit rate is slightly reduced compared to the Hit rate taking all the claims into account. The use of a threshold, either of 2 or 5 claims, considerably increases the Hit rate of the original dataset. Using the POH maximum values and looking at the POH threshold of 50 %, the use of a threshold of 2 claims increases the Hit rate by 0.23, while a threshold of 5 claims increases the Hit rate by 0.29 (Fig. 4.1a). The Hit rate is generally highest using the POH maximum values per postal code area and lowest using the 50<sup>th</sup> percentile per postal code area, even if the Hit rate based on the latter is increased by the use of a threshold of 2 or 5 claims (Fig. 4.1d). With a minimal threshold of 2 claims, the Hit rate always exceeds 0.5.

Generally, the FAR decreases with the POH thresholds (Fig. 4.2). The decrease is stronger above 60 % POH. The use of a threshold of claims slightly increases the FAR, as expected, but this is not desired. For the POH maximum values per postal code area, the use of thresholds of claims increases the FAR by 0.01 to 0.03 below the POH threshold of 60 % and by 0.02 to 0.07 above this value (Fig. 4.2a). Up to the POH threshold of 30 %, the application of thresholds of claims gives the same FAR for a threshold of 2 or 5 claims. Above the POH threshold of 30 %, the difference between the two thresholds of claims is limited to 0.01 to 0.02. The difference between the thresholds of claims is more noticeable when using percentiles of the POH and not the maximum values. For the 50<sup>th</sup> percentile per postal code area, taking all the claims into account gives a  $FAR = 0.86$  for a POH threshold of 80 %. The use of a threshold of 2 claims gives a  $FAR = 0.92$  and a threshold of 5 claims gives a  $FAR = 0.95$  (Fig. 4.2d). The lowest FAR is generally obtained when using the 50<sup>th</sup> percentile of POH values per postal code area. With a threshold of 2 claims minimum, the FAR is never lower than 0.85.

In the original dataset, the Hit rate of the POH decreases more strongly than the FAR (Fig. 4.3). The very high FAR means that there are many *false alarm* cases. The CSI and the HSS are always close to 0 and the Bias is always higher than 1. This is due to the fact that these scores are affected by the high number of *false alarm* cases. The very low HSS means that the detection is more or less similar to the reference level found by chance. The high Bias means that there is a strong overdetection of the hail events.

## MESHHS

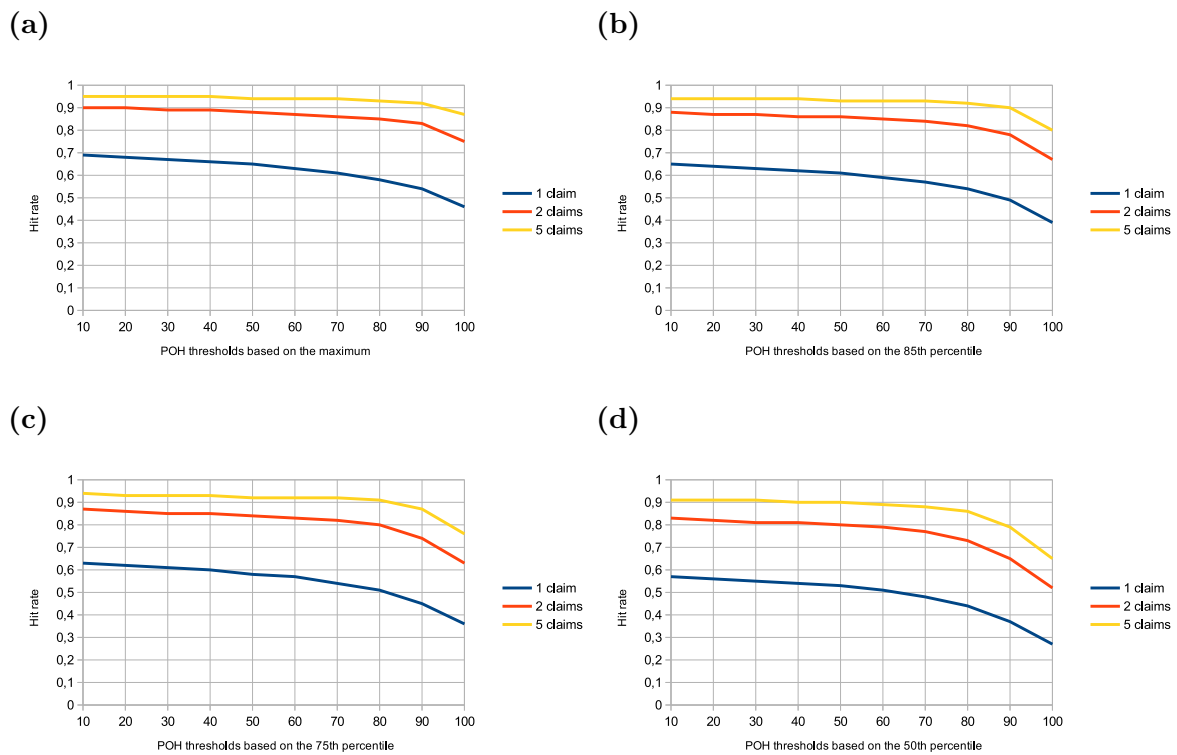
The Hit rate decreases as function of increasing MESHHS thresholds (Fig. 4.4). The decrease of the Hit rate is more or less regular when taking all the claims into account (blue line). Using a threshold of 2 or 5 claims (red and yellow lines) leads to a stronger decrease of the Hit rate for each of the MESHHS thresholds. However, the Hit rate is considerably higher using a threshold of 2 or 5 claims, mainly for the lowest MESHHS thresholds (20-40 mm). Using a threshold of 2 claims, the Hit rate for a MESHHS threshold of 20 mm varies between 0.62 and 0.81 according to the percentile of MESHHS values used. The Hit rate strongly decreases and varies between 0.05 and 0.23 for a MESHHS value of 60 mm. The highest Hit rate is obtained with the MESHHS maximum values (Fig. 4.4a).

With a minimal threshold of 2 claims, the Hit rate is always greater than 0.5 for the MESHHS thresholds of 20 and 30 mm. For higher MESHHS thresholds, the Hit rate strongly decreases, mainly when using percentiles of MESHHS values.

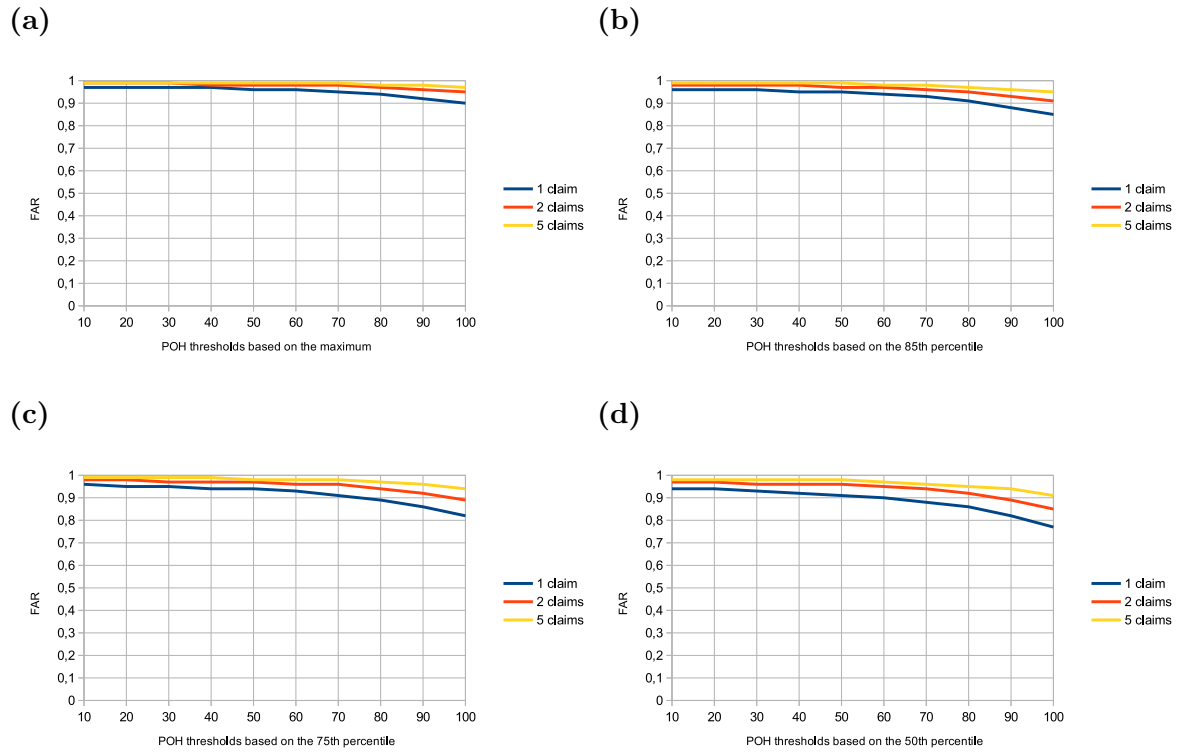
Generally, the FAR decreases quite steadily with the MESHHS thresholds (Fig. 4.5). The use of a threshold of claims slightly increases the FAR (between 0.04 and 0.21), as expected even if this is not desired. The lowest FAR is generally obtained with the use of the 50<sup>th</sup> percentile of MESHHS values. For the 50<sup>th</sup> percentile, taking all the claims into account gives a  $FAR = 0.73$  for a MESHHS threshold of 30 mm. The use of a threshold of 2 claims gives a  $FAR = 0.82$  and a threshold of 5 claims gives a  $FAR = 0.89$  (Fig. 4.5d). With a minimal threshold of 2 claims, the FAR is never lower than 0.65.

In the original dataset, similarly to the POH algorithm, the Hit rate of the MESHHS algorithm decreases more strongly than the FAR. This is not desired as the goal is to have the highest Hit rate and the lowest FAR.

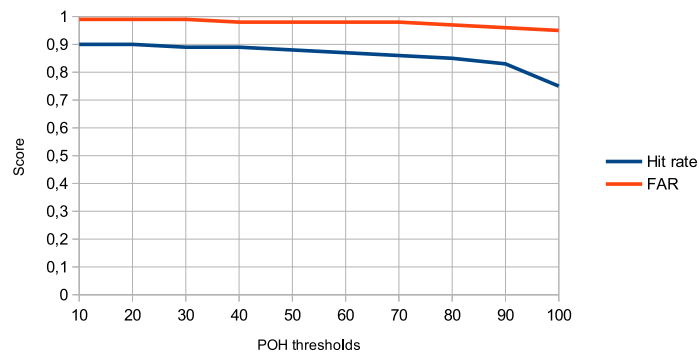
Similarly to the POH algorithm, the very high FAR means that there are many *false alarm* cases. The CSI and the HSS are always close to 0 and the Bias is often higher than 1. This is due to the fact that these scores are affected by the high number of *false alarm* cases. However, the Bias still shows results closer to 1 for the high MESHHS thresholds, mainly when using percentiles. When using the MESHHS maximum values and a threshold of 2 claims, the MESHHS threshold of 60 mm has a  $Bias = 1.64$ , which indicates only a small overdetection of the hail events. When using the 50<sup>th</sup> percentile and a threshold of 2 claims, the MESHHS thresholds of 50 and 60 mm have a Bias lower than 1, indicating an underdetection of the hail events. There is thus generally a strong overdetection of the hail events for the lowest MESHHS thresholds, while the highest MESHHS thresholds provoke an underdetection of the hail events.



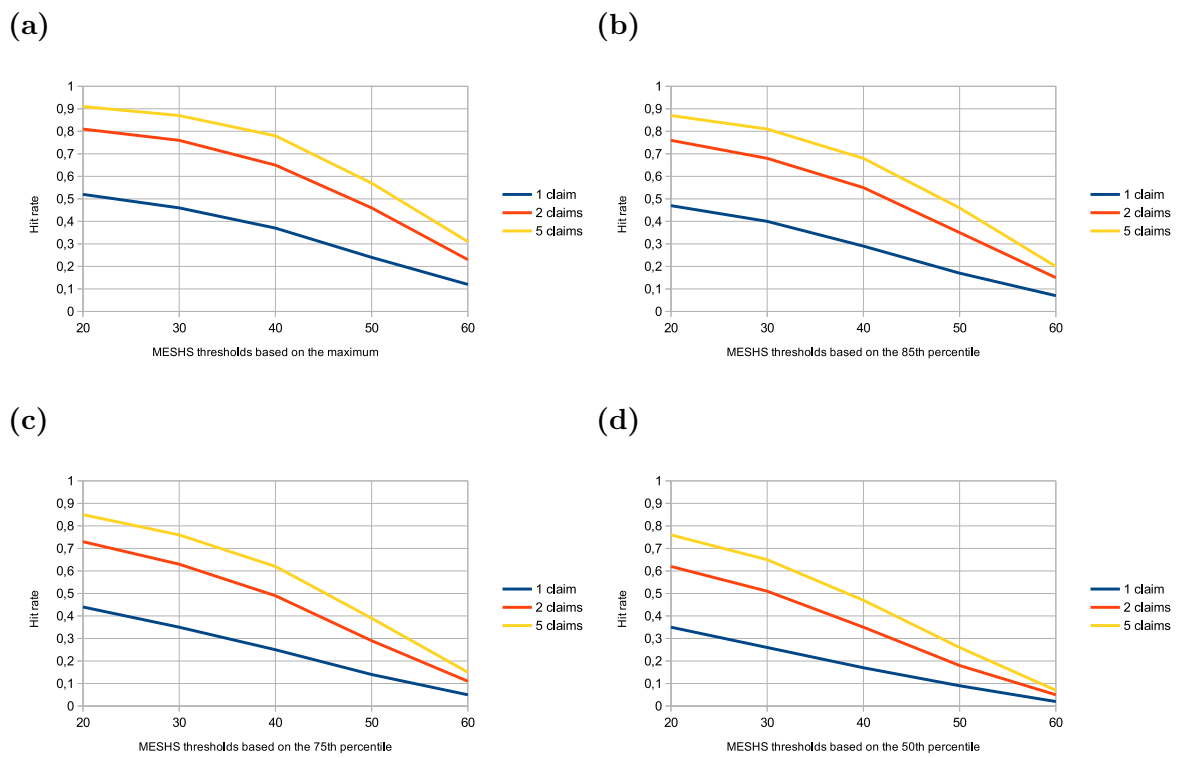
**Figure 4.1:** Hit rate for different POH thresholds for the original dataset, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum (a), on the 85<sup>th</sup> percentile (b), on the 75<sup>th</sup> percentile (c) and on the median (d) within a postal code area.



**Figure 4.2:** FAR for different POH thresholds for the original dataset, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum (a), on the 85<sup>th</sup> percentile (b), on the 75<sup>th</sup> percentile (c) and on the median (d) within a postal code area.

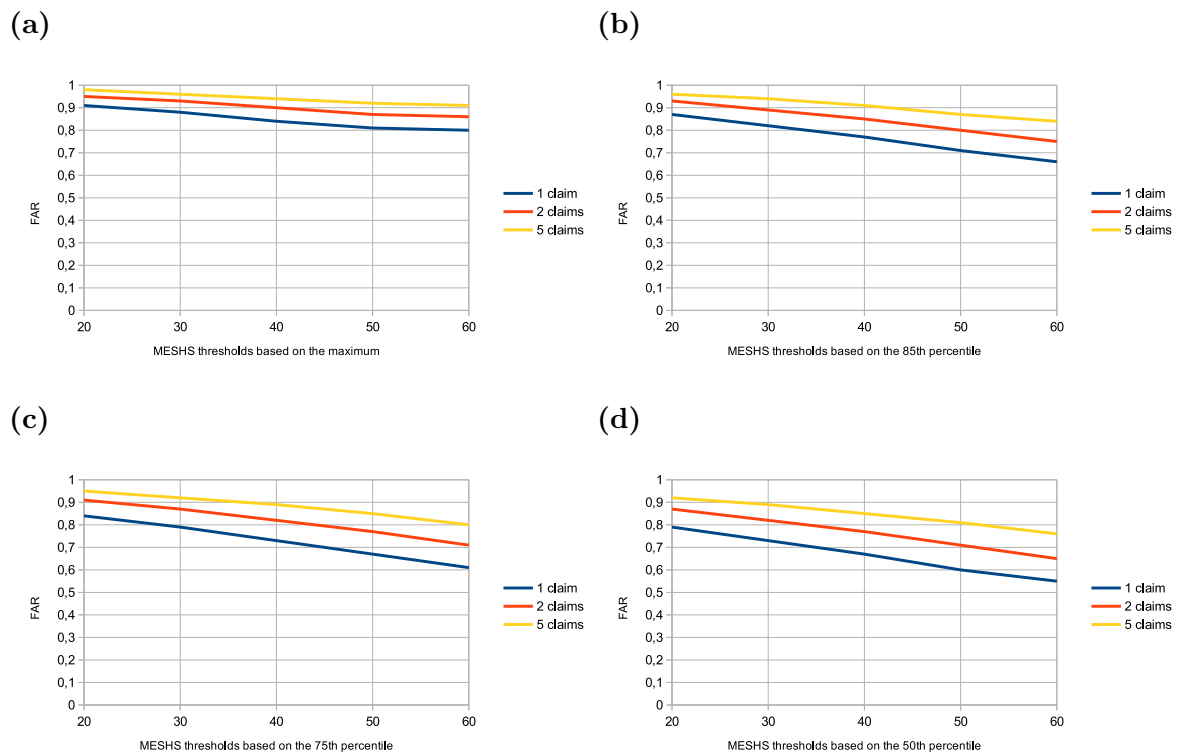


**Figure 4.3:** Hit rate and FAR for different POH thresholds for the original dataset, using a threshold of 2 claims and based on the POH maximum within a postal code area.



**Figure 4.4:** Hit rate for different MESHS thresholds for the original dataset, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHS maximum (a), on the 85<sup>th</sup> percentile (b), on the 75<sup>th</sup> percentile (c) and on the median (d) within a postal code area.



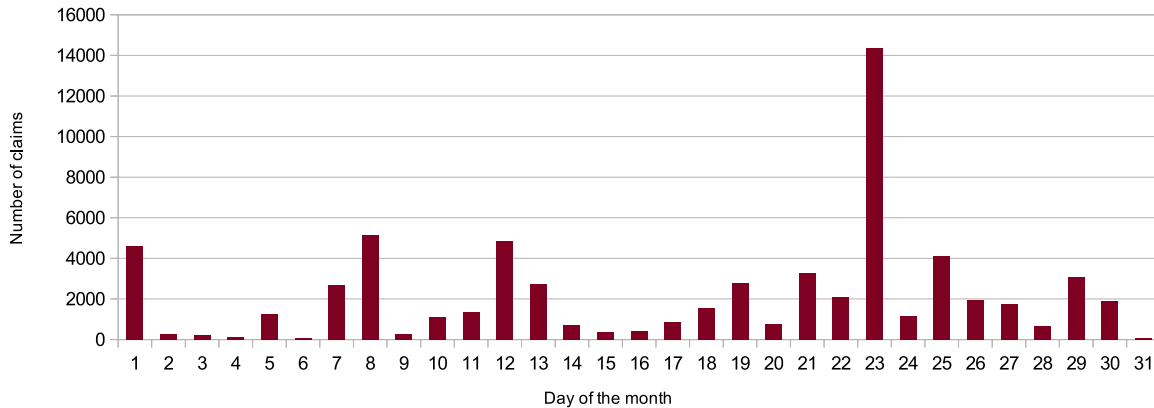


**Figure 4.5:** FAR for different MESHS values for the original dataset, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHS maximum (a), on the 85<sup>th</sup> percentile (b), on the 75<sup>th</sup> percentile (c) and on the median (d) within a postal code area.

### 4.1.2 Dataset with insurance corrections

After the corrections on the insurance dataset (Chapter 3.5.1), 404 hail days and 65'361 claims remained for the time period 2003-2012. Only 1.77 % of the insurance claims have been removed, while more than half of the hail days have been removed. This is due to the presence of many days which had only very few claims. These were sometimes completely removed from the dataset and therefore the day as well.

In his analysis of floods with insurance data, Studer [2013] observed that, until 1993, a majority of the claims were dated on the first of the month. This correlation is important to take into account as it can strongly bias the analysis if it is realised on a daily basis. The present study concerns the years 2003-2012 and such a correlation is thus not expected. Based on the dataset with insurance corrections, there is indeed no correlation between the number of claims and the day of the month (Fig. 4.6). The 23<sup>rd</sup> has a particularly higher amount of claims than the other days but should not influence the results. One specific day, the 23<sup>rd</sup> of July 2009, is mainly responsible for this high amount of claims. It was indeed a particularly extreme hail event, covering a large part of Switzerland.



**Figure 4.6:** Number of claims for each day of the month for the time period 2003-2012, based on the dataset with insurance corrections.

### POH

The Hit rate of the dataset with insurance corrections follows the same pattern as the Hit rate of the original dataset: decrease with increasing POH thresholds, stronger decrease between the POH thresholds of 80 and 100 %, increase with the use of a threshold of 2 or 5 claims and best Hit rate using the POH maximum values (Fig. 4.7). Besides, the Hit rate is increased thanks to the corrections made on the insurance dataset (Chapter 3.5.1). Using the POH maximum values and taking all the claims into account gives an increase of the Hit rate between 0.09 and 0.14 (Fig. 4.8a). Using the POH maximum values and a threshold of 2 claims, the Hit rate is only increased between 0.02 and 0.04 (Fig. 4.8b). With a threshold of 2 claims minimum, the Hit rate is always greater than 0.5.

The FAR obtained from the dataset with insurance corrections is very similar to the FAR

obtained from the original dataset (Fig. 4.9). Based on the POH maximum values and using a threshold of 2 claims, the corrections made on the insurance dataset decrease the FAR between 0.01 and 0.02 only. This is negligible as the FAR remains above 0.8 for any of the percentiles of POH values using a threshold of 2 claims (Fig. 4.10).

The corrections made on the insurance dataset allow to increase the Hit rate of the POH algorithm considerably and only slightly decrease the FAR. The CSI, the Bias and the HSS obtained from the dataset with insurance corrections are similar to the results obtained from the original dataset; the CSI and the HSS are always close to 0 and the Bias is always greater than 1.

## MESHS

Similarly to the original dataset, the Hit rate of the MESHS algorithm for the dataset with insurance correction decreases with increasing MESHS thresholds (Fig. 4.11). The Hit rate follows the same pattern as for the original dataset according to the threshold of claims and the use of percentiles of MESHS values. Looking at the POH maximum values with a threshold of 2 claims, the corrections made on the insurance dataset increase the Hit rate between 0.01 and 0.02 (Fig. 4.12b). The increase is more important taking all the claims into account, mainly for the lowest MESHS thresholds (Fig. 4.12a).

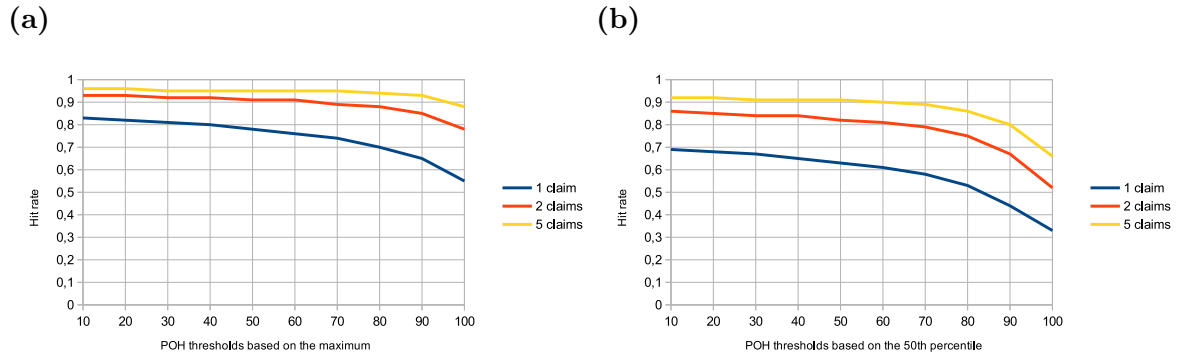
With a threshold of 2 claims minimum, the Hit rate is always greater than 0.5 for the MESHS thresholds of 20 and 30 mm. For higher MESHS thresholds, the Hit rate strongly decreases, mainly when using percentiles of MESHS values.

The FAR of the MESHS algorithm obtained from the dataset with insurance corrections is also very similar to the FAR obtained from the original dataset (Fig. 4.13). The corrections made on the insurance dataset only slightly reduce the FAR for the highest MESHS thresholds based on the MESHS maximum values (Fig. 4.14a). With the use of percentiles of MESHS values, there is almost no decrease of the FAR after the insurance corrections (Fig. 4.14b).

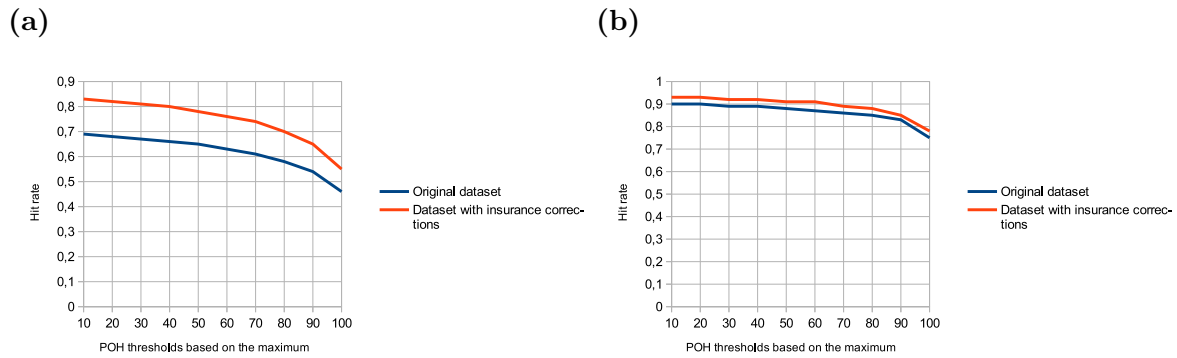
Based on the MESHS maximum values and using a threshold of 2 claims, the FAR of the dataset with insurance corrections remains above 0.8.

Similarly to the original dataset, the Hit rate of the MESHS algorithm decreases more strongly than the FAR as function of the increasing MESHS thresholds. This is not desired as the goal is to have the highest Hit rate and the lowest FAR. The corrections made on the insurance dataset have a smaller impact on the Hit rate and the FAR of the MESHS algorithm than on the POH algorithm.

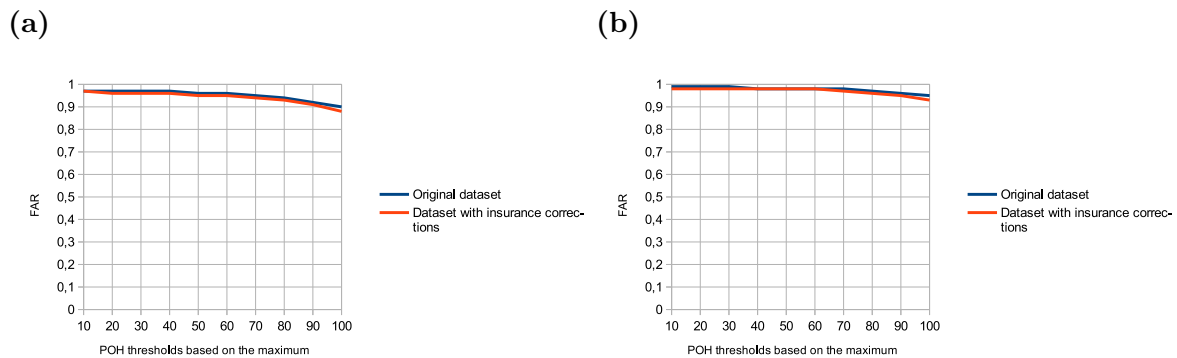
The CSI, the Bias and the HSS obtained from the dataset with insurance corrections are similar to the results obtained from the original dataset.



**Figure 4.7:** Hit rate for different POH thresholds for the dataset with insurance corrections, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum values (a) and on the median (b) within a postal code area.



**Figure 4.8:** Hit rate for different POH thresholds for the original dataset and the dataset with insurance corrections, taking all the claims into account (a) or using a threshold of 2 claims (b). Results based on the POH maximum values within a postal code area.



**Figure 4.9:** FAR for different POH thresholds for the original dataset and the dataset with insurance corrections, taking all the claims into account (a) or using a threshold of 2 claims (b). Results based on the POH maximum values within a postal code area.

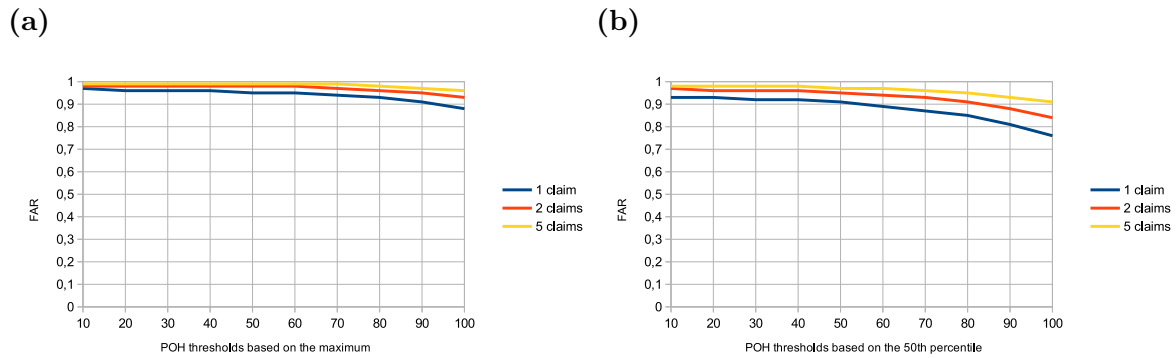


Figure 4.10: FAR for different POH thresholds for the dataset with insurance corrections, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum values (a) and on the median (b) within a postal code area.

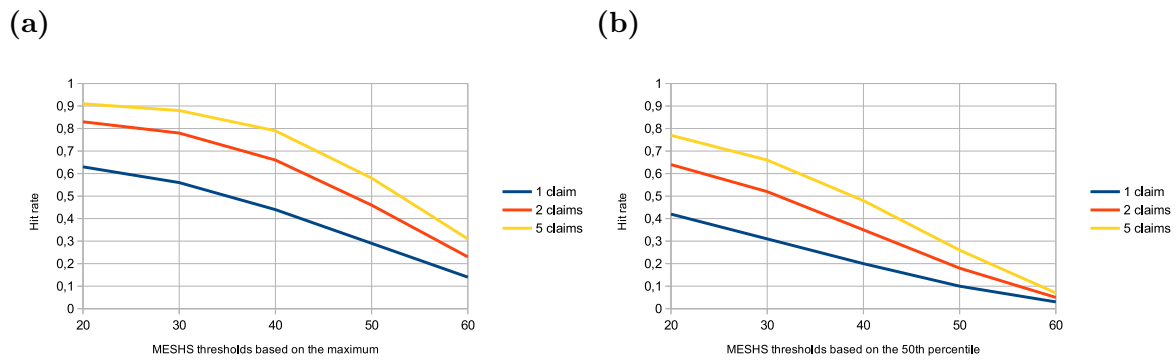


Figure 4.11: Hit rate for different MESHS thresholds for the dataset with insurance corrections, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHS maximum values (a) and on the median (b) within a postal code area.

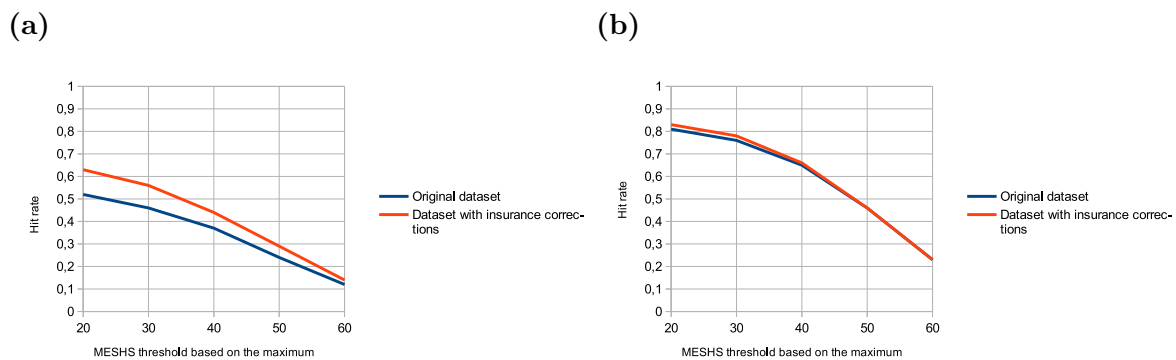
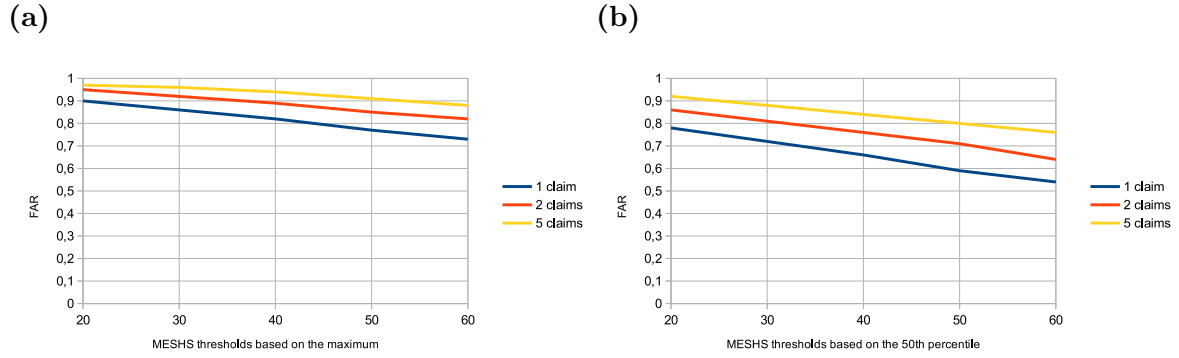
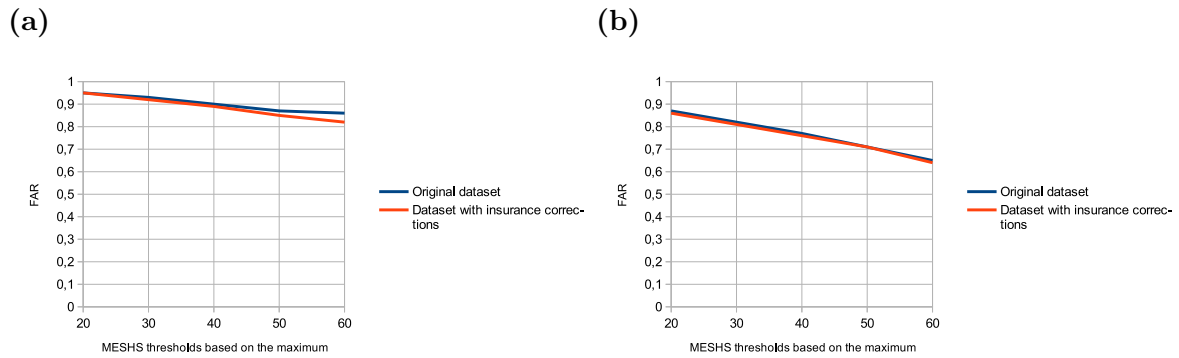


Figure 4.12: Hit rate for different MESHS thresholds for the original dataset and the dataset with insurance corrections, taking all the claims into account (a) or using a threshold of 2 claims (b). Results based on the POH maximum values within a postal code area.



**Figure 4.13:** FAR for different MESH thresholds for the dataset with insurance corrections, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESH maximum values (a) and on the median (b) within a postal code area.



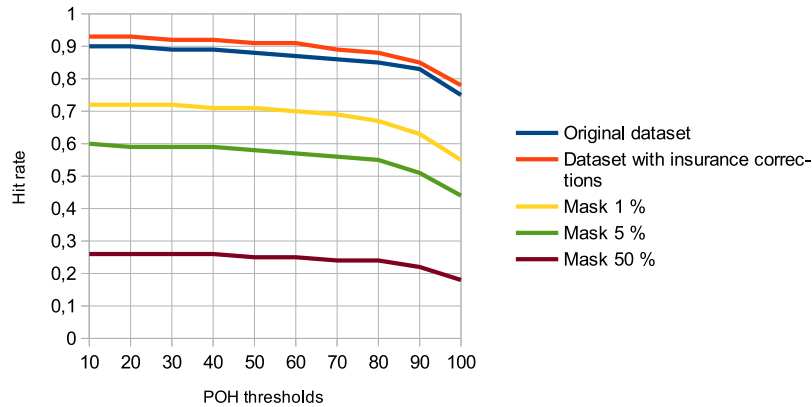
**Figure 4.14:** FAR for different MESH thresholds for the original dataset and the dataset with insurance corrections, using a threshold of 2 claims, based on the MESH maximum values (a) and on the 50<sup>th</sup> percentiles of MESH values (b).

### 4.1.3 Datasets with the masks

#### POH

Similarly to the two previous datasets, the Hit rate of the datasets with the masks decreases as function of increasing POH thresholds. The pattern concerning the thresholds of claims is also the same. However, the Hit rate is much lower than for the previous datasets (Fig. 4.15), mainly for the mask 50 %. This is due to the removal of radar detections on several pixels, while the claims remain at the same location. There are thus more claims considered as *missed* cases, because of the removal of radar detections. Based on the POH maximum values, using a threshold of 2 claims and compared to the dataset with insurance corrections, the Hit rate is reduced by about 0.2 with the mask 1 %, by about 0.3 with the mask 5 % and by about 0.6 with the mask 50 %. The difference of Hit rate between the POH maximum values and percentiles of POH values is much lower than for the previous datasets.

For the mask 1 %, with a threshold of 2 claims minimum, the Hit rate is always greater than 0.4. For the mask 5 %, with a threshold of 2 claims minimum, the Hit rate is always greater than 0.3.



**Figure 4.15:** Hit rate for different POH thresholds for the original dataset, the dataset with insurance corrections and datasets with the masks, using a threshold of 2 claims. Results based on the POH maximum values within a postal code area.

The FAR obtained from the datasets with the masks is very similar to the FAR obtained from the two previous datasets. Based on the POH maximum values and using a threshold of 2 claims, the use of the mask 1 % decreases the FAR by only 0.01 compared to the dataset with insurance corrections. With the mask 5 %, the FAR is reduced by only 0.02. Moreover, when using the 50<sup>th</sup> percentile of POH values with the masks, the FAR is even higher than for the dataset with insurance corrections.

The decrease obtained with the masks based on POH maximum values is mostly negligible as the FAR remains above 0.8 for any of the percentiles applied on the POH values using a threshold of 2 claims. The use of the masks strongly reduces the Hit rate of the POH algorithm, while the FAR is only slightly reduced.

The CSI and the HSS obtained from the datasets with the masks are similar to the results obtained from the two previous datasets; they are always close to 0. The Bias is mostly higher than 1, indicating an over-detection of the hail events, but approaches the value of 1 for high POH values when using the mask 50 % and mostly with the 50<sup>th</sup> percentile.

## MESHS

Similarly to the two previous datasets, the Hit rate of the MESHS algorithm for the datasets with the masks decreases with increasing MESHS thresholds. The Hit rate follows the same pattern as for the other datasets according to the threshold of claims and the use of percentiles of MESHS values. However, the Hit rate is much lower than previously, reaching 0.7 at the maximum, based on the MESHS maximum values with a threshold of 5 claims for the mask 1 %. For the mask 50 %, the Hit rate does not exceed 0.24. For the masks 1 % and 5 %, the Hit rate is still considerably higher for the lowest MESHS thresholds than for the highest MESHS thresholds. With a minimal threshold of 2 claims, the Hit rate of the MESHS algorithm varies between 0.04 (mask 50 %) and 0.62 (mask 1 %).

The FAR of the MESHS algorithm obtained from the datasets with the masks is very similar to the FAR obtained from the previous datasets. With the use of a mask, the FAR decreases more strongly with increasing MESHS thresholds, and mainly by increasing the percentage of traffic area used for the mask. The FAR is generally slightly higher with the masks 1 % and 5 %. With a threshold of 2 claims minimum, the FAR of the MESHS algorithm remains above 0.5.

Similarly to the previous datasets, the Hit rate of the MESHS algorithm decreases more strongly than the FAR with increasing MESHS thresholds. This is not desired as the goal is to have the highest Hit rate and the lowest FAR. The use of the masks 1 % and 5 % reduces the Hit rate and slightly increases the FAR of the MESHS algorithm, but the goal was to observe the opposite.

The CSI and the HSS obtained from the datasets with the masks are similar to the results obtained from the other datasets; they are always close to 0. The Bias is mostly higher than 1 for the masks 1 % and 5 % for MESHS thresholds up to 40 mm, indicating an over-detection of the hail events.

### 4.1.4 Datasets “removed postal codes without claims”

The Hit rate for the datasets “removed postal codes without claims” is not presented as it is not affected by the removal of the postal code areas. For both algorithms (POH and MESHS), the decrease of the FAR with the removal of postal code areas without any claims in 10 years is negligible. Using a threshold of 2 claims, the FAR remains above 0.8 for the POH algorithm and above 0.6 for the MESHS algorithm.

### 4.1.5 Results per postal code areas

The Hit rate and the FAR are calculated for the POH thresholds of 50 and 90 % and for the MESHS thresholds of 20 and 40 mm, based on the dataset with insurance corrections.



---

However, the number of insurance data is too small to have consistent and reliable results, as there are often less than 10 claims per postal code area. Indeed, only one *hit* case in the postal code area is sufficient to retrieve a Hit rate of 1. If there is additionally one *missed* case in the postal code area, the Hit rate is immediately divided by two and equal to 0.5. Similarly to the previous datasets, the FAR is also affected by the lack of insurance data.

#### 4.1.6 Summary of the most important scores

The Hit rate of each dataset based on the POH/MESHS maximum values and using a threshold of 2 claims are presented in Table 4.1 for the POH algorithm and in Table 4.2 for the MESHS algorithm. It is possible to see that the datasets “removed postal codes without claims” produce the same results as the ones on which they are based (dataset with insurance corrections and dataset mask 1 %). The dataset with insurance corrections has the best Hit rate and is generally 3 % higher than the original dataset.

POH threshold \ Dataset	Original dataset	Dataset with insurance corrections	Dataset with the mask 1 %	Dataset with the mask 5 %	Dataset with the mask 50 %	Dataset with the insurance correction "removed postal codes without claims"	Dataset with the mask 1 % removed postal codes without claims
10	0.90	0.93	0.72	0.60	0.26	0.93	0.72
20	0.90	0.93	0.72	0.59	0.26	0.93	0.72
30	0.89	0.92	0.72	0.59	0.26	0.92	0.72
40	0.89	0.92	0.71	0.59	0.26	0.92	0.71
50	0.88	0.91	0.71	0.58	0.25	0.91	0.71
60	0.87	0.91	0.70	0.57	0.25	0.91	0.70
70	0.86	0.89	0.69	0.56	0.24	0.89	0.69
80	0.85	0.88	0.67	0.55	0.24	0.88	0.67
90	0.83	0.85	0.63	0.51	0.22	0.85	0.63
100	0.75	0.78	0.55	0.44	0.18	0.78	0.55

**Table 4.1:** Hit rate of the POH algorithm for the different datasets. Results based on the POH maximum values, using a threshold of 2 claims.

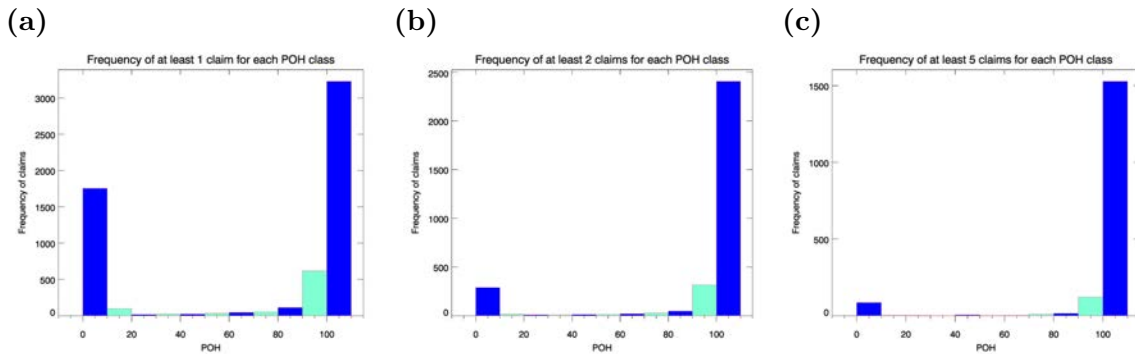
POH threshold \ Dataset	Original dataset	Dataset with insurance corrections	Dataset with the mask 1 %	Dataset with the mask 5 %	Dataset with the mask 50 %	Dataset with the insurance correction "removed postal codes without claims"	Dataset with the mask 1 % removed postal codes without claims
20	0.81	0.83	0.62	0.49	0.21	0.83	0.62
30	0.76	0.78	0.56	0.44	0.18	0.78	0.56
40	0.65	0.66	0.45	0.35	0.13	0.66	0.45
50	0.46	0.46	0.30	0.22	0.08	0.46	0.30
60	0.23	0.23	0.14	0.11	0.04	0.23	0.14

**Table 4.2:** Hit rate of the POH algorithm for the different datasets. Results based on the MESHS maximum values, using a threshold of 2 claims.

## 4.2 Frequency of claims per POH/MESHS values

### POH

The histograms (Fig. 4.16) are based on the dataset with insurance corrections as it produced the best Hit rate. They show the frequency of claims for each of the ten POH classes for the whole time period. The histogram taking all the claims into account shows a high frequency of claims for POH values between 0 and 9 % (Fig. 4.16a). By increasing the threshold of claims to 2 or 5 claims (Fig. 4.16b and 4.16c), this frequency decreases considerably, going from more than 1700 occurrences of claims to less than 400. The frequency in the other POH classes decreases as well when increasing the threshold of claims, but more moderately. Most of the occurrences of claims occur for 100 % POH for all the thresholds of claims.



**Figure 4.16:** Histograms showing the frequency of claims for each of the POH classes, using the dataset with insurance corrections. The frequency is shown taking all the claims into account (a), using a threshold of 2 and 5 claims ((b) and (c)).

As can be seen on Fig. 4.17, the cumulated claims for each POH class follows the same pattern as the frequency, that is the frequency of claims increases with the POH values, except for the first POH class. Most of the claims occur with a POH of 100 % and many claims are present for POH values between 0 and 9 %.

### MESHS

The histograms for MESHS (Fig. 4.18) are also based on the dataset with insurance corrections. They show the frequency of claims for each of the six MESHS classes. The first class considers the values between 0 and 19 mm, while the other classes consider each only 10 mm. This is due to the algorithm which retrieves values starting from 20 mm. When using a threshold of 2 or 5 claims, in a general way, the frequency of claims increases with the MESHS values. But similarly to the histograms of the POH algorithm, a high frequency of claims is observed in the first MESHS class, particularly for the histogram taking all the claims occurrences into account (Fig. 4.18a). This high frequency decreases by more than five times with the use of a threshold of 2 claims (Fig. 4.18b). This shows that many single claims are corresponding to MESHS values lower than 20 mm. However, using a threshold of claims also decreases the frequency considerably in the other MESHS classes. Most of the classes

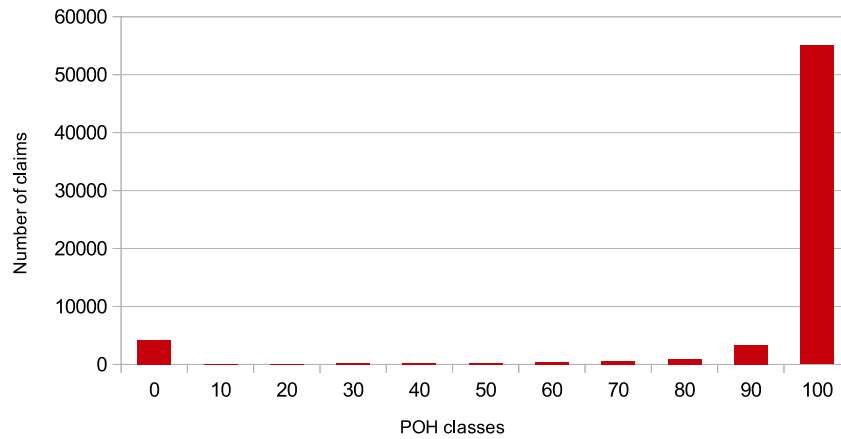


Figure 4.17: Total number of claims for each POH class for the period 2003-2012, based on the dataset with insurance corrections.

that have a strong decrease in the frequency of claims are the lower classes (20-30 mm). This means that the higher classes also have the most important damage, that is at least 2 or 5 claims for the same event in the same postal code area. That is also why, when taking all the claims into account, the lowest MESHS values (20-50 mm) have a higher frequency of claims than the MESHS class 60 mm. Besides, relatively to the lowest MESHS classes, the frequency of claims of the highest MESHS classes (50-60 mm) increases with a threshold of 2 or 5 claims. Compared to the POH frequencies, the claims occurrences are more equally distributed between the MESHS classes. Using a threshold of 2 claims, there is a noticeable difference between the MESHS classes 30-39 mm and 40-49 mm, with higher frequencies starting from the MESHS class 40-49 mm. This class is also the one where the frequency is higher than the one of the MESHS class 0-19 mm.

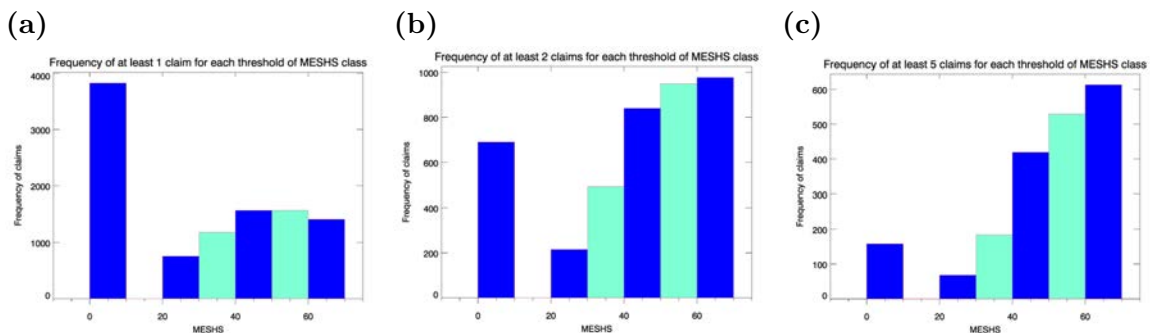
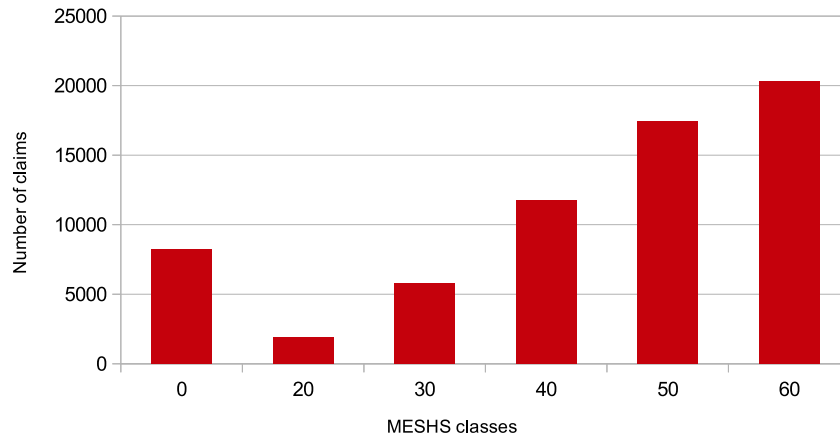


Figure 4.18: Histograms showing the frequency of claims for each of the MESHS classes, using the dataset with insurance corrections. The frequency is shown taking account all the claims into (a), using a threshold of 2 and 5 claims ((b) and (c)).

The pattern of the total number of claims for each MESHS class is smoother than the one of the frequencies (Fig. 4.19). There is no important difference between any of the MESHS classes, except between the first one and the second one. This is still due to the high amount

of claims for the MESH values between 0 and 19 mm and it is consistent with the results shown by the histograms. Here again, the MESH class 40-49 mm is the one where more claims start to be found than in the MESH class 0-19 mm.



**Figure 4.19:** Total number of claims for each MESH class for the period 2003-2012, based on the dataset with insurance corrections.

## 5 Discussion

Radar-based hail detection has several advantages compared to measurement stations. For example, it has a global coverage of the country, mainly in the low land, and high spatial and temporal resolutions. This is very important for rare short living meteorological events, such as hailstorms. Difficulties arise with the validation of the results, given the lack of ground-based observations. The present study aims at verifying the POH and MESHS radar-based hail detection algorithms. The results obtained are discussed in this section and compared with other similar studies.

### 5.1 Verification scores

The following analysis focuses on the Hit rate and the FAR, as they are the most representative scores assessing the quality of the radar detection and taking into account the most important parameters. Moreover, if one of these two scores is strongly biased, such as the FAR in the present study, it will inevitably affect the scores ensuing from them (CSI, Bias, HSS). The latter will thus mostly be biased.

#### 5.1.1 POH

##### Original dataset

Several adjustments have been made on the original radar and insurance datasets in order to improve the quality of the validation, which is mainly based on the Hit rate and the FAR. The latter was indeed very high for the original dataset due to several reasons that are explained more in detail in Chapter 5.2.

The use of thresholds of claims to select the postal code areas containing a minimum number of claims (1, 2 or 5) considerably increases the Hit rate and slightly increases the FAR. The difference of Hit rate between the thresholds of 2 and 5 claims is low, while it is more important between the thresholds of 1 and 2 claims (Fig. 4.1). This indicates that many single claims in the original dataset are probably incorrect (Chapter 2.2.1). They correspond to *missed* cases (Fig. 3.3) and their removal increases the Hit rate by making the *hit* cases higher in proportion. These single claims might also be correct but without corresponding to any radar detection. This is the case if there is strong wind shear drifting the hailstones and leading to damage farther away from the radar detection. Several methods can be used to reduce the impact of this phenomenon on the results; using bigger regions, using a threshold

of claims or using buffers on the radar data, similarly to the method used by W. Schmid and H. H. Schiesser [1st European Hail Workshop, 2014] who horizontally extended the footprint of the hail cloud by 1 to 5 km. The method using a threshold of claims is probably the easiest one to apply and already allows to remove part of this uncertainty. This method is particularly useful if no corrections are done on the insurance dataset, as it can be seen on Fig. 4.1. The use of thresholds of claims is only favourable to the Hit rate. The latter increased between 0.21 and 0.29 for the original dataset based on the POH maximum values. For the FAR, the use of percentiles of POH values is the method which produces the best results if no corrections or adjustments, such as the use of a mask or a threshold of claims, are made on the dataset.

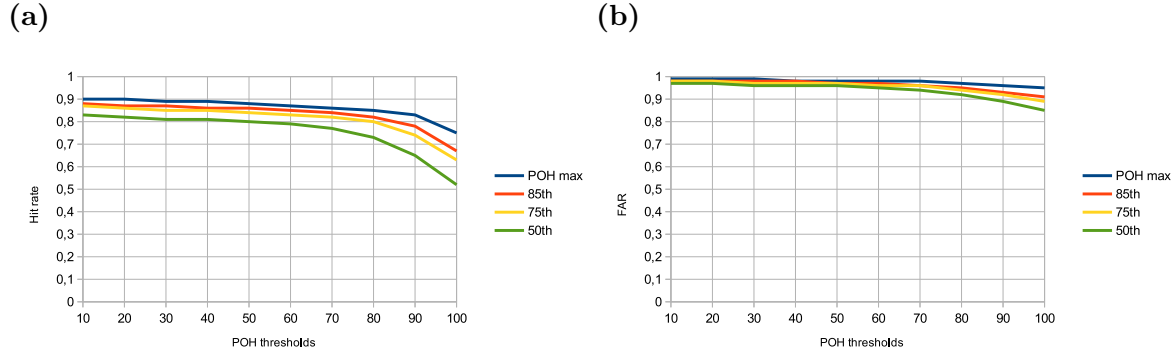
The impact of thresholds of claims or of percentiles of POH values on the Hit rate and the FAR follows the same pattern for every dataset. That is why recommendations concerning the use of these two variables are already made here, in order to focus on the most relevant results in the following of the discussion.

The choice of the threshold of claims for the POH algorithm verification should thus be based on the goal to achieve; the maximisation of the Hit rate, of the FAR or a balance between both. The Hit rate is best with a threshold of 5 claims, while the FAR is best when taking all the claims into account. If the goal is to find a balance, the threshold of 2 claims should be used, as it limits the increase of the FAR without reducing too much the Hit rate. A threshold of 2 claims is preferred anyway as it limits the risk of removing too much correct data or keeping too much incorrect data. As shown in Chapter 4, the FAR remained high for any of the adjustments or corrections applied on the original dataset. Besides, the FAR is close to 100 % and even taking all the claims into account does not allow to strongly reduce it. The FAR cannot be considered as reliable as it is not possible to know if the very high results are due to a lack of insurance data on the ground or if there is indeed a strong over-detection by the algorithm. That is why the present study uses the Hit rate as reference for the verification of the POH algorithm. However, the threshold of 2 claims is still recommended to be used for the analysis instead of the threshold of 5 claims as the risk of removing correct data is lower and it still produces a high Hit rate. Moreover, the difference of Hit rate between a thresholds of 2 and 5 claims is relatively small; between 0.05 and 0.12 using the POH maximum values on the original dataset.

Similarly to the threshold of claims, the choice of the percentile of POH values to use for the algorithm verification should be based on the goal to achieve; the maximisation of the Hit rate, of the FAR or a balance between both. The Hit rate is best with the POH maximum values (Fig. 5.1a), while the FAR is best when using the 50<sup>th</sup> percentile of POH values (Fig. 5.1b). Using the POH maximum values gives a higher Hit rate because more pixels have a radar detection compared to when using percentiles of POH values and there are therefore less *missed* cases. The 50<sup>th</sup> percentile of POH values gives a better FAR because there is less radar detection as the extreme values of the postal code area are weighted. However, the FAR is only slightly reduced by the use of percentiles. Moreover, as previously said, the Hit rate is mainly used as reference for the verification of the POH algorithm. That is why it is recommended to mainly use the POH maximum values for the analysis. It is also possible to use a lower percentile on POH values to weight the extreme POH values, but the difference of Hit rate remains slight. The Hit rate and FAR have additionally been calculated with the four different percentiles of POH values to see the variation according to the adjustments



made on the original dataset.



**Figure 5.1:** Comparison of the Hit rate (a) and of the FAR (b) of the POH algorithm for the POH maximum, the 85<sup>th</sup>, the 75<sup>th</sup> and the 50<sup>th</sup> percentiles of POH values for the original dataset using a threshold of 2 claims.

The high Hit rate found for almost all the datasets means that, generally, the POH algorithm detects most of the hail events that occurred. The Hit rate decreasing with the POH thresholds means that there is an increase of the *missed* events. This is consistent with the fact that not all the claims occur for 100 % POH (Fig. 4.7a). However, the slight decrease of the Hit rate shows that most of the claims take place with high POH values. A stronger decrease of the Hit rate occurs for the POH threshold of 100 %. The reason is that a considerable amount of claims considered as *hit* for a POH detection of 90 % are then considered as *missed* for the threshold of 100 %. This means that many claims correspond to a radar detection of 90 %. However, this does not give any information on the number of *hit* cases for a POH value of 100 %. As shown on Fig. 4.17, out of the POH class 0-9 %, a majority of the *hit* cases occur for the POH classes of 90 and 100 %. The stronger decrease of Hit rate for the POH threshold of 100 % gives a first indication concerning the POH threshold above which important damage on cars occurs, that is here 90 %.

### Dataset with insurance corrections

The insurance corrections made on the original dataset reduce the uncertainty due to single claims without any radar detection. As expected, the Hit rate increases considerably when taking all the claims into account (Fig. 4.8a). The increase of the Hit rate is lower when using a threshold of 2 or 5 claims (Fig. 4.8b). This confirms the presence of single claims that are very likely incorrect. The FAR is only very slightly reduced as only few claims are moved by the correction on pixels with a radar detection that did not already have a claim. The Hit rate is even better with the application of a threshold of 2 claims on the corrected dataset. The single claims that are removed by the threshold of claims are probably correct, but only one claim in a postal code area is too little to be considered as reliable. Given the higher Hit rate obtained using the dataset with insurance corrections, it seems to be worth correcting the insurance data. This allows to strongly reduce the uncertainty caused by errors in the date and location of the claims. It is thus recommended to use this dataset as basis for any further analysis on the verification of the POH algorithm. The use of a threshold of

claims does not strongly improve the Hit rate or the FAR but, for the reasons cited before, a threshold of 2 claims is recommended nonetheless.

### Datasets with the masks

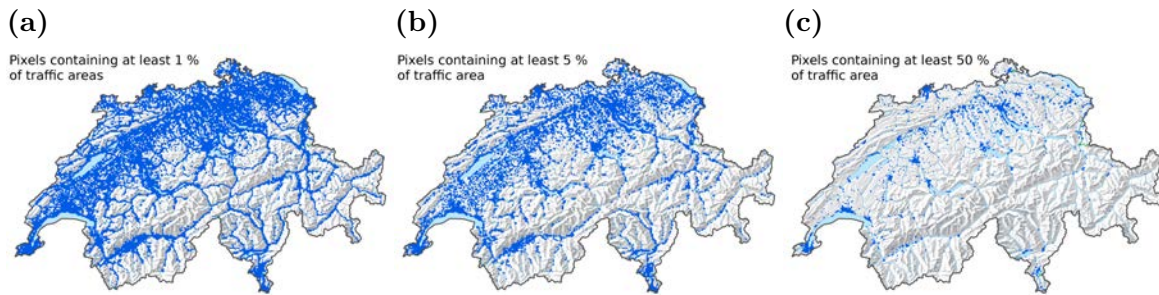
The use of a mask on the radar dataset was mainly meant to reduce the FAR. However, the latter is only slightly reduced with the use of the masks, even with the mask 50 %. Moreover, the Hit rate strongly decreases when using them. This is not desired, as a strong decrease of the FAR is expected for only a small decrease of the Hit rate. The errors in the mask presented in Chapter 3.5.2 do not fully explain the results obtained. As can be seen on Fig. 5.2a, the mask 1 % does not strongly reduce the coverage of the radar detection as areas with more traffic generally correspond to areas where hail is more frequent. Besides, when using the POH maximum values, only one pixel in a postal code area is enough to give its value to the whole postal code area. However, the use of a percentile of POH values does not strongly decrease the FAR either. Better results were mainly expected with the masks 5 % and 50 % (Fig. 5.2b and 5.2c). But once again, the FAR does not considerably decrease, while the Hit rate, in the contrary, strongly decreases. Several reasons can explain the high FAR over urban areas, even with the use of a mask:

- A possibility is that many of the hail events in these areas are in reality mainly composed of grapel and thus do not create any damage to cars.
- It is also possible that few or no insured cars were present in many pixels remaining in the mask at the moment of the hail event.
- A reason can be the presence of an uncertainty between the radar detection and hail on the ground. As it will be presented in Chapter 5.2, the wind shear can influence the location on the ground of the hailfall, which can strongly differ from the radar detection in altitude.
- It might simply be due to the POH algorithm strongly overestimating the presence of hail.

However, the present study does not allow to validate or reject any of these hypothesis.

The strong decrease of the Hit rate when using a mask is due to the increase of the *missed* cases by removing correct radar detections corresponding to claims. Indeed, if one or several claims are present in a postal code area where the POH values of the pixels have been set to zero due to the mask, the claims will remain there and be counted as *missed* cases, decreasing the Hit rate.

The mask used is thus not adapted to the datasets used in the present study. More information concerning the traffic area and modelisation would be needed in order to improve the mask. In their study, Kunz and Kugel [2014] used as well a mask, but considering buildings, agriculture areas, water areas and vegetation/forest areas. However, the results that they obtained were not much different from the ones obtained in the present study.



**Figure 5.2:** Masks applied on the dataset with insurance corrections and showing the pixels with at least 1 % (a), 5 % (b) and 50 % (c) of traffic area per pixel for Switzerland [graphics: L. Thomi].

### Datasets “removed postal codes without claims”

The dataset giving the best results is generally the one with insurance corrections “removed postal codes without claims”. However, the difference of Hit rate and FAR with the dataset with insurance corrections is close to zero. As the removal of these postal code areas does not strongly increase the Hit rate or reduce the FAR, the dataset with insurance corrections can be used as reference for the verification of the POH algorithm to reduce the complexity of the analysis. Removing postal code areas that do not contain any claims is still meaningful, but the reasons for these missing claims should be further investigated to understand why it is the case and what it implies.

### CSI, Bias, HSS

For any of the datasets, POH values, thresholds of claims and percentiles of POH values, the CSI and the HSS are very low (CSI between 0.02 and 0.07, and HSS between 0.02 and 0.12 for the dataset with insurance corrections, using the POH maximum values and a threshold of 2 claims) and the Bias is high (between 11.34 and 60.52 for the dataset with insurance corrections, using the POH maximum values and a threshold of 2 claims). This is mainly due to the very high number of *false alarm* cases, variable taken into account in the three scores. Having obtained a high FAR before, such results were expected for the CSI, the Bias and the HSS. In accordance with the high FAR and very good Hit rate obtained, the large number of *false alarm* cases can account for the high CSI. The HSS is often close to zero and this means that the detection is just a little bit better than the random detection obtained by chance. Here again the score is influenced by the large amount of *false alarm* cases. However, the HSS never goes below zero and this means that the detection is never worse than the reference detection. The high Bias indicates a strong overdetection of hail events by the radar, once again mainly due to the large amount of *false alarm* cases. The uncertainties related to *false alarm* cases should first be reduced, before giving too much importance to these scores.

### POH threshold for important damage

The following analysis is mainly based on the dataset with insurance corrections, using a threshold of 2 claims and the POH maximum values.

One goal of the study is to determine above which POH threshold important damage occurs. The histograms (Fig. 4.16 and 4.17) show that, for a threshold of 2 claims, more than three quarters of the claims occurrences occur for POH values between 90 and 100 %. Besides, the results obtained for the Hit rate show that a majority of the damage occurs above a POH threshold of 90 % (Fig. 4.7). However, it is important to consider that cars are less vulnerable than other objects, such as crops. The latter are usually more affected by hail as small hailstones (< 2 cm) can lead to strong damage. That is why the POH threshold must be adapted according to the purpose. For the damage on cars, a lower POH threshold of 90 % can be used as representative of important damage, as most of the claims correspond to POH values higher than 90 %. This threshold should probably be lower for crops and higher for buildings.

### 5.1.2 MESHS

#### Original dataset

The same adjustments (the use of a mask on radar data, corrections on the insurance data, use of percentiles of MESHS values, use of thresholds of claims) as for the POH algorithm have been applied on the original dataset in order to improve the results, as the FAR was also particularly high. For the MESHS algorithm, the use of thresholds of claims allows to considerably increase the Hit rate (Fig. 4.4). This is due to the fact that the MESHS algorithm produces values only for the most severe hailstorms. And thus, many single claims that would correspond to a low POH value do not correspond anymore to any radar detection for the MESHS algorithm. The number of claims remaining the same for both algorithms, this increases the amount of *missed* cases and decreases the Hit rate. It explains the strong decrease of the Hit rate with the MESHS thresholds (see for example Fig. 4.11a); hailstones of 4 to 6 cm are more unusual than smaller hailstones. Thus, the number of *missed* cases increases with the MESHS thresholds, being considered as *hit* cases for the lower thresholds. The decrease of the Hit rate is also due to the fact that high MESHS values (40-60 mm) are more rarely detected than lower MESHS values. For the original dataset based on the MESHS maximum values and using a threshold of 2 claims, there is a Hit rate of only 0.23 for the MESHS threshold of 60 mm. With the assumption that all the *missed* cases are due to the *hit* cases for lower MESHS thresholds, this result means that 23 % of the claims occurrences occur for a MESHS of 60 mm. The Hit rate of 0.81 for the MESHS threshold of 20 mm means that about 20 % of the claims occurrences occur for MESHS values lower than 20 mm. This is consistent with the assumption that important damage occur for MESHS values above 20 mm.

The MESHS algorithm only indicates that if its value is equal or higher than 20 mm, a POH detection is expected. Apart from this, it provides information on the size of the hailstones, so it is difficult to relate it to the number of claims.

Using the original dataset, the FAR for the MESHS algorithm is considerably increased by the use of a threshold of claims, especially with 5 claims. Similarly to the POH algorithm, the use of thresholds of claims is only favourable to the Hit rate. For the FAR, the use of percentiles of MESHS values is the most favourable method if no corrections or adjustments are made. Taking all the claims into account gives the lowest results for the FAR, but the

use of a threshold of 2 or 5 claims only slightly increases the FAR. The reasons are the same as for the POH algorithm. Generally, the FAR is lower than for the POH algorithm. The explanation is the same as for the lower Hit rate, that is the increase of *missed* cases with the MESHS algorithm. This shows that the overdetection of hail events is lower when focusing on the strongest ones, which is the case of the MESHS algorithm. However, the FAR is still very high ( $> 0.5$ ). This is mainly due to several uncertainties concerning the FAR (Chapter 5.2), such as the lack of ground truth, the hail falling several meters away from the radar detection due to a strong wind shear, the uncertain location of the cars or the method used to link the radar pixels with the postal code areas. The verification of the MESHS algorithm is thus also mainly based on the Hit rate. That is why a threshold of 2 claims is recommended to be used as well for the verification of the MESHS algorithm.

The choice of the percentile of MESHS values to use is made similarly to the POH algorithm. As the Hit rate is mainly used as reference for the verification of the MESHS algorithm, the MESHS maximum values are recommended to be used for analysis.

### Datasets with insurance corrections

Similarly to the POH algorithm, the insurance corrections made on the original dataset reduce the uncertainty due to the single claims without any radar detection. As expected, the Hit rate increases considerably when taking all the claims into account. The impact is almost non-existent when using a threshold of 2 claims (Fig. 4.12b). This confirms as well the presence of single claims that are probably incorrect. The FAR is also reduced but only very slightly as only few claims are moved by the correction on pixels with a radar detection that did not already have a claim. The decrease of the FAR is stronger for the MESHS thresholds of 50 and 60 mm. This means that, with the correction of the insurance data, more claims are moved from pixels with no MESHS value to pixels with high MESHS values than to pixels with low MESHS values.

The dataset giving the best Hit rate for the MESHS algorithm, and which can easily be generalized for further analysis, is the one with insurance corrections (0.23-0.83). It is therefore recommended to use this dataset as a basis for any further analysis on the verification of the MESHS algorithm. The use of a threshold of claims does not strongly improve the Hit rate or the FAR but, for the reasons cited before (limitation of the increase of the FAR without reducing too much the Hit rate), a threshold of 2 claims is recommended nonetheless.

### Datasets with the masks

The best FAR is obtained with the dataset with the mask 50 %, using the 50<sup>th</sup> percentile of MESHS values. The behaviour of the Hit rate and the FAR according to the mask follows the same pattern as for the POH algorithm; a strong decrease of the Hit rate and only a slight decrease for the FAR. The reasons for this behaviour are the same as for the POH algorithm: the low reduction of the radar coverage by the mask 1 % with the remaining of the claims, the possible presence of graupel, the uncertainty concerning the location of the cars or the difference between hail actually occurring on the ground and the radar detection.

### Datasets “removed postal codes without claims”

Similarly to the POH algorithm, the removal of postal code areas without any claims in 10 years does not improve the FAR considerably. The dataset giving the best results is generally the one with insurance corrections “removed postal codes without claims”. However, in order to make the study more generalized, the dataset with insurance corrections only is used as reference. Removing postal code areas that do not contain any claims is not wrong, but the reasons for these missing claims should be more investigated to understand why it is the case and what it implies.

### CSI, Bias, HSS

For any of the datasets, MESHS values, thresholds of claims and percentiles of MESHS values, the CSI and the HSS are very low (CSI between 0.05 and 0.11, and HSS between 0.09 and 0.20 for the dataset with insurance corrections, using the MESHS maximum values and a threshold of 2 claims). Similarly to the POH algorithm, this is very much due to the high number of *false alarm* cases, but also to the higher number of *missed* cases. The analysis is thus the same as for the POH algorithm: the HSS close to zero means that the detection is just a little bit better than the random detection obtained by chance. Concerning the Bias, it varies more considerably according to the dataset used, MESHS values, thresholds of claims and percentiles of MESHS values. Focusing on the dataset with insurance corrections and using the MESHS maximum values, the Bias is generally higher than 1, indicating an over-detection of the hail events. The Bias is closer to 1 when taking all the claims into account. When using the 50<sup>th</sup> percentile of MESHS values, the Bias is lower for lower MESHS values which corresponds mainly to an under-detection of the hail events. Using a threshold of 2 claims, the detection is the less biased for the MESHS values between 40 and 50 mm. There is an over-detection for lower MESHS values and an under-detection for higher MESHS values. This under-detection is consistent with the MESHS algorithm focusing on the most severe hailstorms.

### MESHS threshold for important damage

The following analysis is based mainly on the dataset with insurance corrections, using a threshold of 2 claims and the MESHS maximum values.

The Hit rate quickly decreases with the MESHS thresholds and is close to zero for a threshold of 60 mm. This means that many claims correspond to *hit* cases for each MESHS threshold and are thus considered as *missed* cases for the higher MESHS thresholds, which decreases the Hit rate. This is due to the fact that there are many more pixels with low MESHS values than pixels with high MESHS values [Luca Nisi, personal communication, August 2014]. To know the approximate “real” Hit rate of high MESHS thresholds, the *hit* cases of the previous MESHS thresholds should be removed in order not to be counted as *missed* cases. And the *missed* cases which correspond to a MESHS value lower than 20 mm should be distributed as *missed* cases among the different thresholds.

One goal of the study is to determine above which MESHS threshold important damage occurs. Based on the results of the Hit rate decreasing more or less regularly with the

MESHS values, it is possible to say that important car damage occurs from MESHS values of 20 mm already. This is also consistent with the results obtained for the frequency of claims related to MESHS classes (Fig. 4.18). It is possible to see that a relatively high number of claims occurrences occur for each of the classes. When using a threshold of 2 claims, about 1400 claims occurrences occur for the MESHS classes between 0 and 30 mm, while the MESHS classes between 40 and 60 mm have almost twice more claims occurrences, about 2800 occurrences. The MESHS class of 40 mm has more than one third of additional claims compared to the MESHS class of 20 mm. Moreover, the MESHS classes of 40 to 60 mm also have a high cumulated number of claims (Fig. 4.19). It is thus recommended to use the MESHS threshold of 40 mm as lower threshold for the most important damage, but taking into account that about half of the claims occurrences still occurs for lower MESHS thresholds. Similarly to the POH algorithm, the MESHS threshold of 40 mm for the most important damage is probably lower for crops and higher for buildings.

These results are consistent with the statement of M. Imhof [1<sup>st</sup> European Hail Workshop, 2014], saying that a damage provoked by hailstones smaller than 20 mm is usually considered as an aesthetic damage, while a damage larger than 20 mm is usually considered as a functional damage.

### 5.1.3 Comparison with other studies

The results obtained in the present study are mostly consistent with the results from previous studies that made verifications of radar-based hail detection algorithms. Most of them focused on the verification of the POH algorithm as it is much more difficult to have a consistent dataset of observations on the ground of the size of the hailstones.

In their study, Kunz and Kugel [2014] used building insurance data. Better results can be expected as the mobility of the damaged object is non-existent, but buildings are also usually less vulnerable to small hailstones. Similarly to the present study, they indeed obtained a very high Hit rate (0.95) but were also confronted to a high FAR ( $> 0.8$ ) for the POH algorithm. They compared different algorithms for the hail detection and all of them substantially overestimated hail occurrence by up to 80 %. The POH algorithm obtained the highest number of *false alarm* cases. The problem of overestimation of the hail events for all of the algorithms seems to come from the verification process using damage data that contains many uncertainties and unknowns (Chapter 5.2). Moreover, Brimelow and Reuter [2006] put forward that the FAR tends to increase with the increase of the area of study. Skripniková [2013] found similar patterns for the POH, with the Hit rate and the FAR decreasing with increasing POH thresholds. The Hit rate decreases more strongly than the FAR that remains higher than 0.4. Through case studies, Kunz and Kugel [2014] obtained similar results for the FAR of the POH algorithm. However, these case studies focus on severe events and it is predictable that the results will be better. This still shows that it is possible to have a lower FAR when there are less uncertainties. In their study, Delobbe et al. [2005] did not even try to calculate the FAR as it would not have been reliable without any assumption on the fraction of hail events that were not reported. For the POH algorithm, Betschart and Hering [2012] found a Hit rate of 0.97 and a FAR of 0.19 over Switzerland. However, they did the analysis only for the years 2009-2011 and weighted the number of reports in the Hit rate and the size of the area in the FAR. For the MESHS algorithm, they found a Hit rate of 0.83 and

a FAR of 0.6.

In their study over Switzerland, Betschart and Hering [2012] concluded that the POH and MESHS algorithms are particularly good at detecting hail events, even mild ones, with a CSI (Critical Success Index) of 0.78 for POH and of 0.79 for MESHS. However, MESHS has a quite low correlation with the ground-truth hail size data. Moreover, POH tends to overestimate the occurrence of hail, while MESHS tends to underestimate it. Skripniková [2013] found a CSI maximum of about 0.3 for a POH equivalent to 100 %, while Kunz and Kugel [2014] only reached a CSI slightly higher than 0.15 for POH values higher than 80 %. This is higher than the CSI obtained in the present study for the dataset with insurance corrections, using a threshold of 2 claims, where it did not exceed 0.15 either for the POH or the MESHS algorithm. The HSS found by Kunz and Kugel [2014] did not exceed 0.3. This is still higher than the HSS found in the present study which did not exceed 0.2.

Concerning the POH threshold above which important damage are expected to occur, Delobbe et al. [2005] found that all hail events with hailstones larger than 2 cm were detected with a POH value larger or equal to 80 %. This value is consistent with the previous analysis of the present study. However, with the results of this one, the POH threshold for important damage could be raised to 90 %.

Based on the results of these different studies, it is possible to see that the results obtained in the present study are consistent with the ones previously found. They can thus be considered as reliable and be used for further analysis.

## 5.2 Discussion concerning the high FAR

A particularly high FAR has been highlighted throughout the present study. Several reasons can explain it and should be taken into account before drawing any conclusions on the results. These reasons are presented in this chapter.

One reason for the high FAR is that there is an uncertainty between the radar detection and hail on the ground. Hail can indeed fall several meters to kilometers away from the radar detection if strong wind shear is present, as this is often the case in hailstorms. This leads to *missed* events where hail did actually fall and provoke damage, and to *false alarms* where the hail is detected by the radar and expected to take place but was not observed on the ground. Moreover, the hailstones can melt on the way down if they are too small [Mahoney et al., 2012] and also increase the *false alarm* cases.

One of the main uncertainty that probably increases the FAR is the lack of information concerning the location of the cars. The latter are indeed moving and it is difficult to know in which proportion the claims are really well located according to the hail event. Moreover, as it has been demonstrated with the use of the mask, it is very difficult to know the areas where cars can be expected or not.

As previously said, the lack of ground-truth verification data is also strongly responsible for the high FAR. It is therefore not possible to know if the high FAR is high due to a bad detection of the radar algorithms or if it is because nobody was there to attest the occurrence of the hail event. Moreover, the observations on the ground based on insurance data depend on the vulnerability of the cars. Indeed, they are generally not affected by hailstones smaller



than 2 cm [M. Imhof, 1<sup>st</sup> European Hail Workshop, 2014]. As the POH algorithm does not give any information concerning the size of the hailstones, a POH of 100 % does not necessarily correspond to a damage on the ground if the size of the hailstones is too small [Kunz and Kugel, 2014].

The low resolution of the insurance data is also to be taken into account. The radar detection is at a spatial resolution of the pixel and the insurance data at a spatial resolution of the postal code area. The size of the postal code areas produces an uncertainty. Some are big and can be spread over an area with meteorological conditions on the ground varying a lot. As already said, one pixel with a high radar value is enough to consider the whole postal code as having this value. However, the use of percentiles of POH/MESHS values did not strongly improve the FAR. It would have been the case if many cases corresponded to this situation. This aspect explains only a small part of the high FAR.

The method used to link the radar pixels with the postal code areas might also bring an uncertainty increasing the FAR. Indeed, many pixels cover several postal code areas. If these pixels have a high radar detection and are in reality *false alarm* cases, the number of *false alarm* cases will be multiplied by the number of postal code areas having these pixels.

A last reason is the modification of the radar signal by several factors, such as the signal attenuation in heavy rain or the beam shielding by the horizon and obstacles [Austin, 1987; Delrieu et al., 1991; Germann et al., 2006; Joss and Waldvogel, 1990; Pellarin et al., 2002; Yuter, 2002]. These modifications occur in the atmosphere and might reduce or change the radar signal according to the composition of the atmosphere. The orography can also slightly influence the radar detection in mountainous areas.

All these factors brought together can indeed explain a large part of the high FAR. Most of them would need to be investigated more in depth in order to strongly reduce the FAR. The different possibilities will be discussed in the Outlook (Chapter 6.1).

As the MESHS focuses only on the most severe hailstorms, it could be interesting to make the following hypothesis. The lowest FAR is obtained with MESHS 60 mm. With the dataset with insurance corrections, using the MESHS maximum values and a threshold of 2 claims,  $FAR = 0.73$ . Given the many uncertainties affecting the FAR, this result could be considered as the best result that can be achieved. The uncertainties can thus be assumed to be equivalent to 70 % of the result obtained for the Hit rate and the FAR. With this hypothesis, the FAR of the POH algorithm would be reduced to 0.23-0.28 for the dataset with insurance corrections, using the POH maximum values and a threshold of 2 claims. These values would be equivalent to the removal of most of the uncertainties affecting the FAR. The new FAR of the POH would thus still show an overdetection by the radar algorithm of the hail events, but in much lower proportions. Given the high Hit rate obtained with the POH algorithm, this slight overdetection could mean that an adaptation of the algorithm could still be tested in order to reduce the *false alarm* cases and reduce the overdetection.

## 6 Conclusion

The present study aimed at verifying the quality of the two radar-based hail detection algorithms POH and MESHS of the MeteoSwiss radar network with the help of insurance loss data for the years 2003-2012. Different methods have been used in order to improve the datasets and to remove part of the uncertainties (Chapter 3). Some corrections have been made on the insurance dataset to remove incorrect data due to the date or location of the claims. A GIS-mask has also been used to reduce the coverage of the radar data to areas where cars could be most expected. Besides, some postal code areas that did not have any claims in 10 years have been removed from the dataset to reduce the FAR. In order to weight the extreme high values of the POH and MESHS algorithms, percentiles (50<sup>th</sup>, 75<sup>th</sup>, 85<sup>th</sup>) have been used per postal code area. Finally, thresholds of claims have been applied on the datasets to remove incorrect claims. The main results and conclusions arising from the analysis in the present study are presented below.

Among the adjustments applied to the original dataset, not all of them brought better results. The dataset which brings the best results is the one with insurance corrections (Chapter 3.5.1). To create this dataset, insurance claims that seemed to be incorrect (without any radar detection in the surroundings) have been removed or moved in time if they corresponded to a radar detection in the surrounding days at the same location.

The maximum radar value per postal code area was generally used. The use of percentiles per postal code area to weight the extreme POH/MESHS values does not strongly differ from the results with the POH/MESHS maximum values. The uncertainty due to the extreme POH/MESHS maximum values is therefore low.

The use of a threshold of 2 claims is useful to reduce the uncertainty due to inaccurate insurance data. This threshold of 2 claims considers single claims in a postal code area as incorrect.

The Hit rate gives accurate and high results. Based on the dataset with insurance corrections, using the POH/MESHS maximum values and a threshold of 2 claims, the Hit rate of the POH algorithm is between 0.78 and 0.93, and between 0.23 and 0.83 for the MESHS algorithm. Besides, the high FAR still contains too many uncertainties to be reliable (Chapter 5.2). The Hit rate should thus first be used for verification studies of radar-based hail detection algorithms. The high Hit rate means that the POH and MESHS algorithms are suited to detect hail events occurring on the ground.

Based on the results obtained with the histograms showing the frequency of claims per POH values (Fig. 4.16) and on the number of claims per postal code area (Fig. 4.17), it can be deduced that important damage on cars occurs for POH values equal to or larger than 90

%. More claims occurrences indeed occur for such values than for lower values. The Hit rate obtained from the dataset with insurance corrections confirms this statement.

The assessment of the quality of the MESHS algorithm is more difficult, as no indication concerning the size of the hailstones is given in the insurance hail reports. The relation between the radar data and insurance reports is thus only indicative. However, it can still be deduced from the Hit rate (Fig. 4.11) and from the histograms (Fig. 4.18) that most of the damage occurs for MESHS values greater than or equal to 40 mm.

The FAR is lower for the MESHS algorithm than for the POH algorithm, as the MESHS algorithm detects severe hailstorms with hailstones larger than 2 cm only and reduces the number of *false alarm* cases.

## 6.1 Outlook

The present study can be used as a basis for further research on radar-based hail detection verification and hail warning. Many other approaches can be used to reduce the remaining uncertainties and improve the verification of the POH and MESHS algorithms. Some of the next possible steps are presented below.

In the present study, the function linking the radar and insurance data together is based on the postal code areas present in each pixel. This method is problematic as it takes into account several times the same pixels. It would be interesting to try using another approach. A possibility would be to spread the pixels in the postal code areas according to their centroids. The pixels would be linked to the postal code area where their centroids are located. This would prevent the multiple counting of several pixels but would also cause other problems. The postal code areas that are too small to contain any pixel centroid would not be taken into account in the analysis. A comparison between the two methods would be interesting and would allow to see which method contains the smallest uncertainties.

The current mask used on the radar data does not improve the results. The goal of the mask is to remove areas where cars are not expected to be present. It would thus be interesting to create a mask taking into account areas where cars are expected, similarly to the one used in the present study, and to then place the claims of each postal code area on the remaining pixels. This could be done by placing the claims first on the pixels that are expected to contain more cars (motorways, city centers, etc.). This method would reduce the uncertainty due to the location of the cars, but due to the finer resolution, the uncertainty concerning the difference between hail detection and hail on the ground could be increased. As described in Chapter 5.2, hail does not always fall exactly where the radar detects it.

To reduce the uncertainty due to the location of the cars considerably, the use of a traffic model would be necessary. It would allow to know where are the cars at every moment of the day. An analysis according to the time of the day would then be necessary.

It could be useful to find an automated way to integrate a wind parameter to the radar-based hail detection algorithms in order to assess the speed and direction of the wind shear. This would allow to adjust the hail detection to the correct location to know where to expect it on the ground. The main difficulty would be to choose which altitude of wind to integrate in the algorithms.

These data processing steps aim at improving the FAR values. However, the best way to do is probably to improve the ground truth coverage. This would reduce the uncertainty concerning the location of the cars and the one due to the lack of observations due to the absence of people to observe the hail event. This enlargement of the ground truth coverage could be done by using several insurance datasets based, for example, on buildings and crops loss data. Moreover, crops loss data might give more information concerning the small hailstones. The hailpads could also be a useful tool to improve the coverage and automatically register hail.

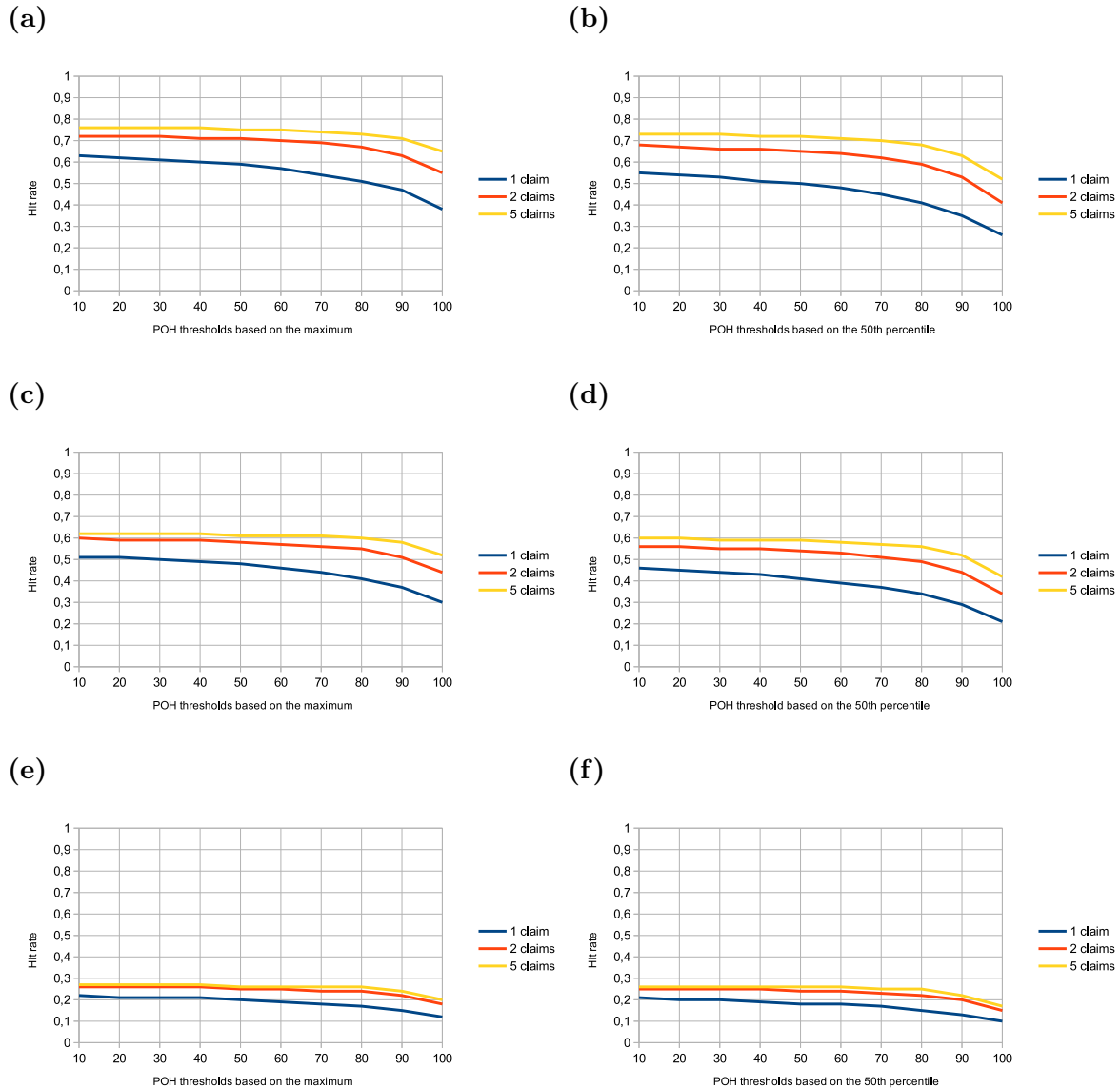
The results obtained in the present study show that the corrections brought by correcting the insurance data were useful. It could thus be interesting to find a way to make these corrections automatically in order to simplify the process.

Based on the verification of the POH and MESHS algorithms, it could be interesting to investigate the relationship between the two algorithms. This might allow to see if there is a correlation between them which would show which size of hailstones can be expected for which POH value.

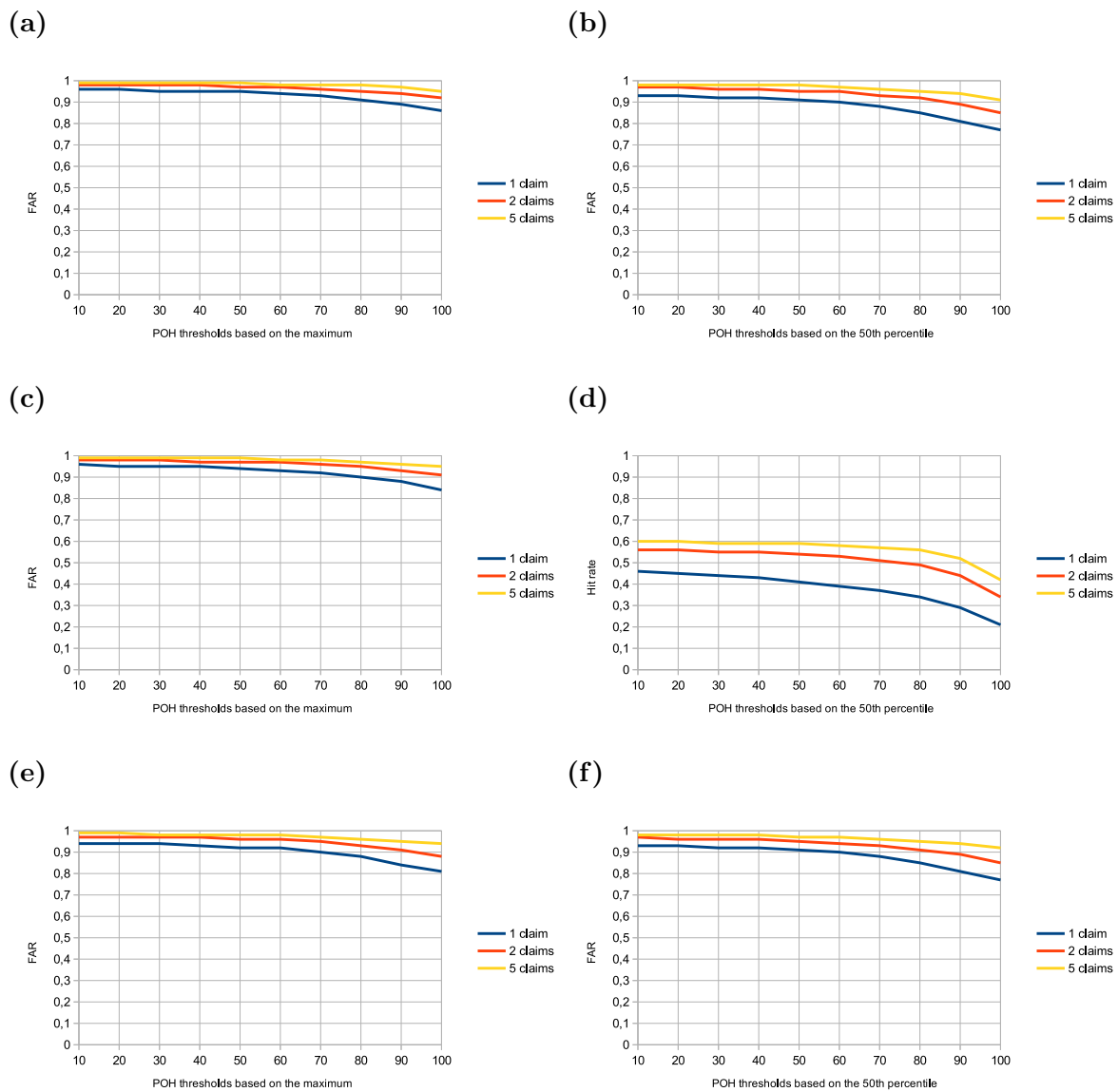
# Appendices

The following results are presented in this chapter as they bring information less relevant for the study.

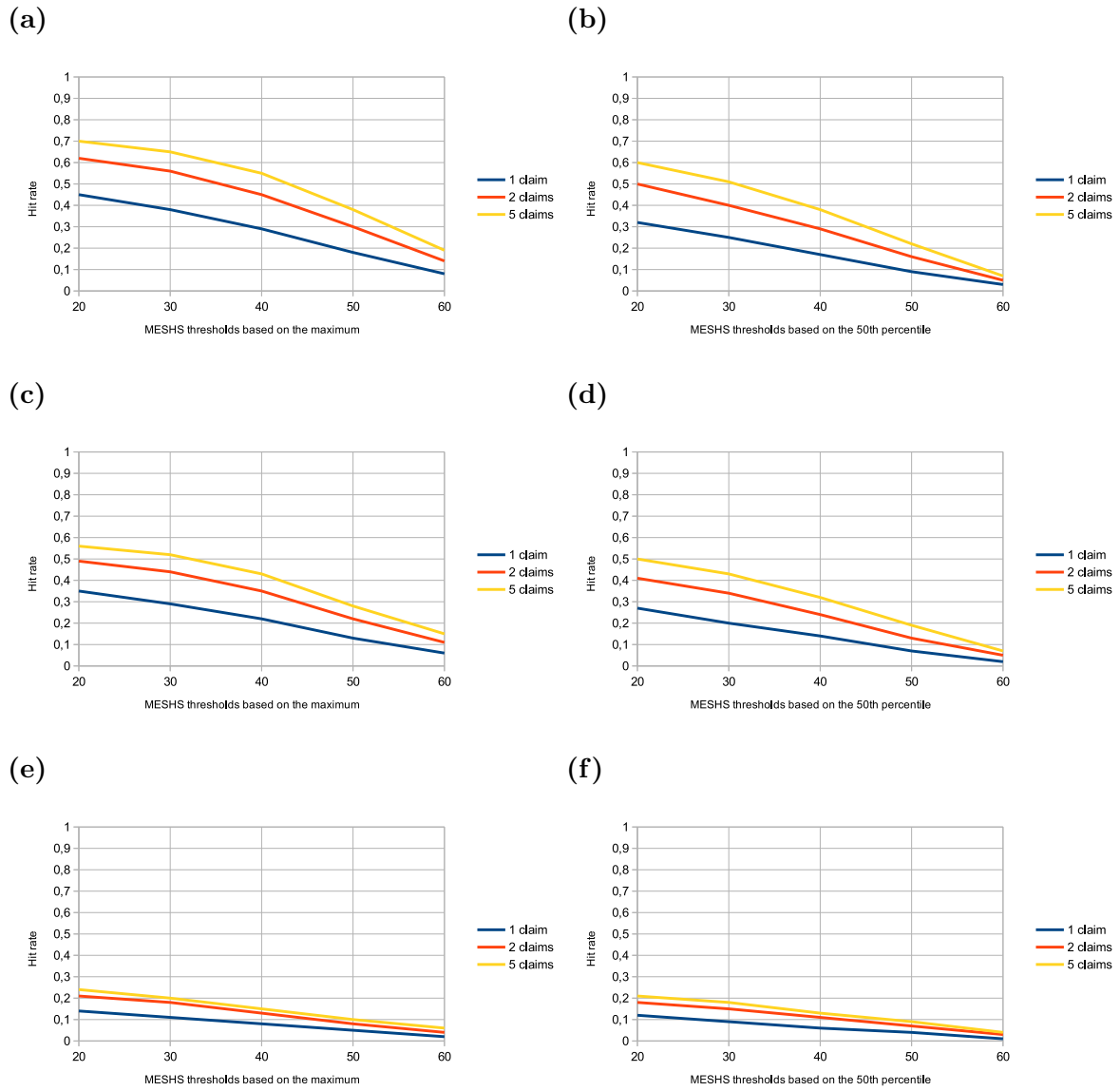
## Results for the datasets with the masks



**Figure 1:** Hit rate for different POH thresholds for the dataset with the masks 1 %, 5 % and 50 %, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum ((a), (c) and (e)) and on the median ((b), (d) and (f)) within a postal code area.

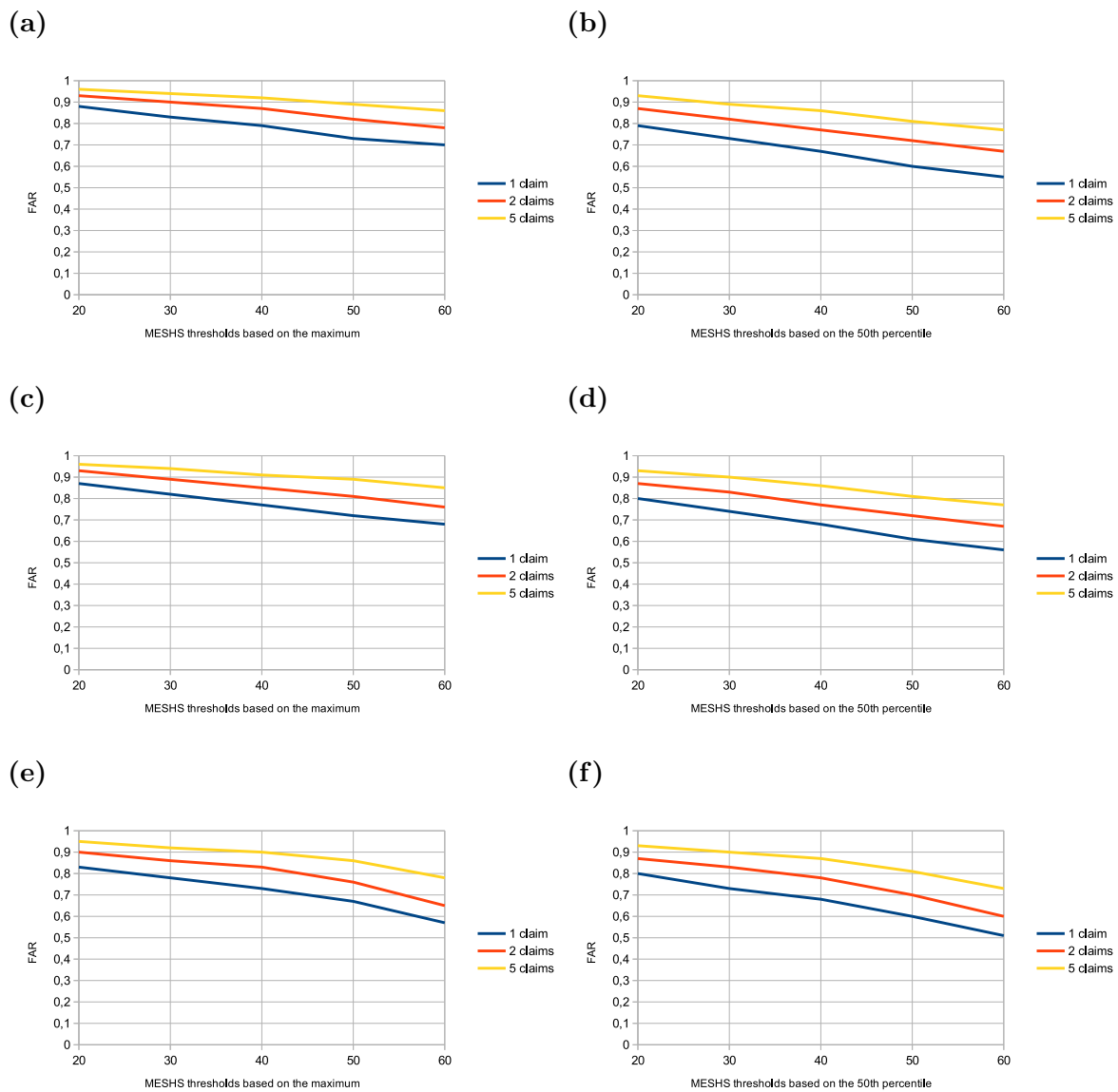


**Figure 2:** FAR for different POH thresholds for the dataset with the masks 1 %, 5 % and 50 %, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum ((a), (c) and (e)) and on the median ((b), (d) and (f)) within a postal code area.



**Figure 3:** Hit rate for different MESHS thresholds for the dataset with the masks 1 %, 5 % and 50 %, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHS maximum ((a), (c) and (e)) and on the median ((b), (d) and (f)) within a postal code area.



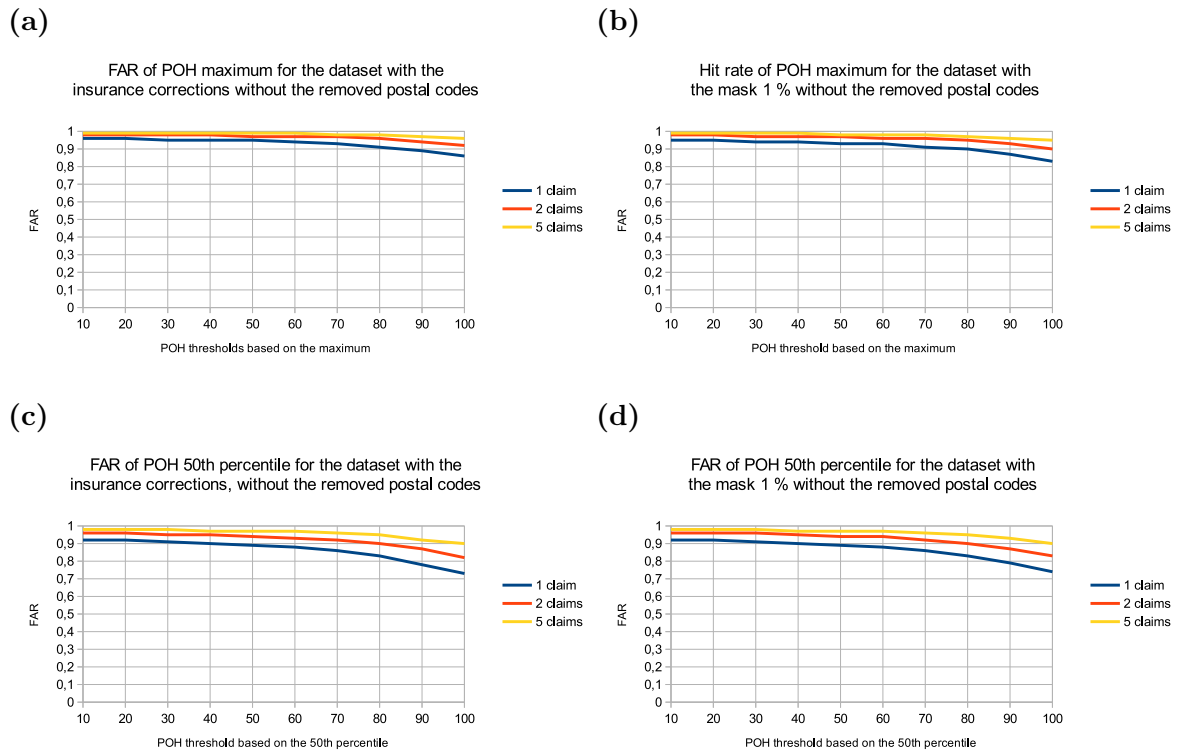


**Figure 4:** FAR for different MESHS thresholds for the dataset with the masks 1 %, 5 % and 50 %, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHS maximum ((a), (c) and (e)) and on the median ((b), (d) and (f)) within a postal code area.

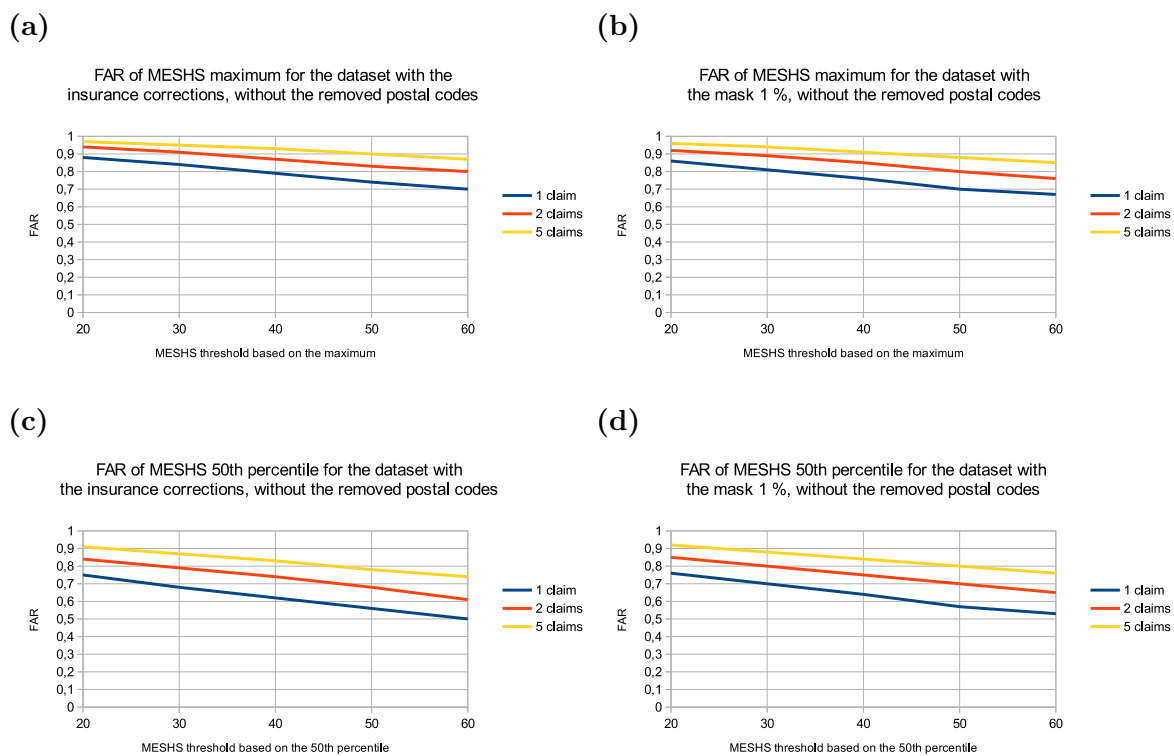
## Datasets “removed postal codes without claims”

The decrease obtained with the removal of postal code areas without any claims in 10 years is negligible as the FAR remains above 0.8 for any of the percentiles of POH values using a threshold of 2 claims (Fig. 5).

Concerning the MESHES algorithm, with the removal of postal code areas without any claims in 10 years, the FAR still remains above 0.6 for any of the percentiles of MESHES values using a threshold of 2 claims (Fig. 6).



**Figure 5:** FAR for different POH thresholds for the dataset with insurance corrections “removed postal codes without claims” ((a) and (c)) and for the dataset with the mask 1 % “removed postal codes without claims” ((b) and (d)). A distinction is made for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum ((a) and (b)) and on the median ((d) and (c)) within a postal code area.



**Figure 6:** FAR for different MESHS thresholds for the dataset with insurance corrections “removed postal codes without claims” ((a) and (c)) and for the dataset with the mask 1 % “removed postal codes without claims” ((b) and (d)). A distinction is made for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHS maximum ((a) and (b)) and on the median ((c) and (d)) within a postal code area.

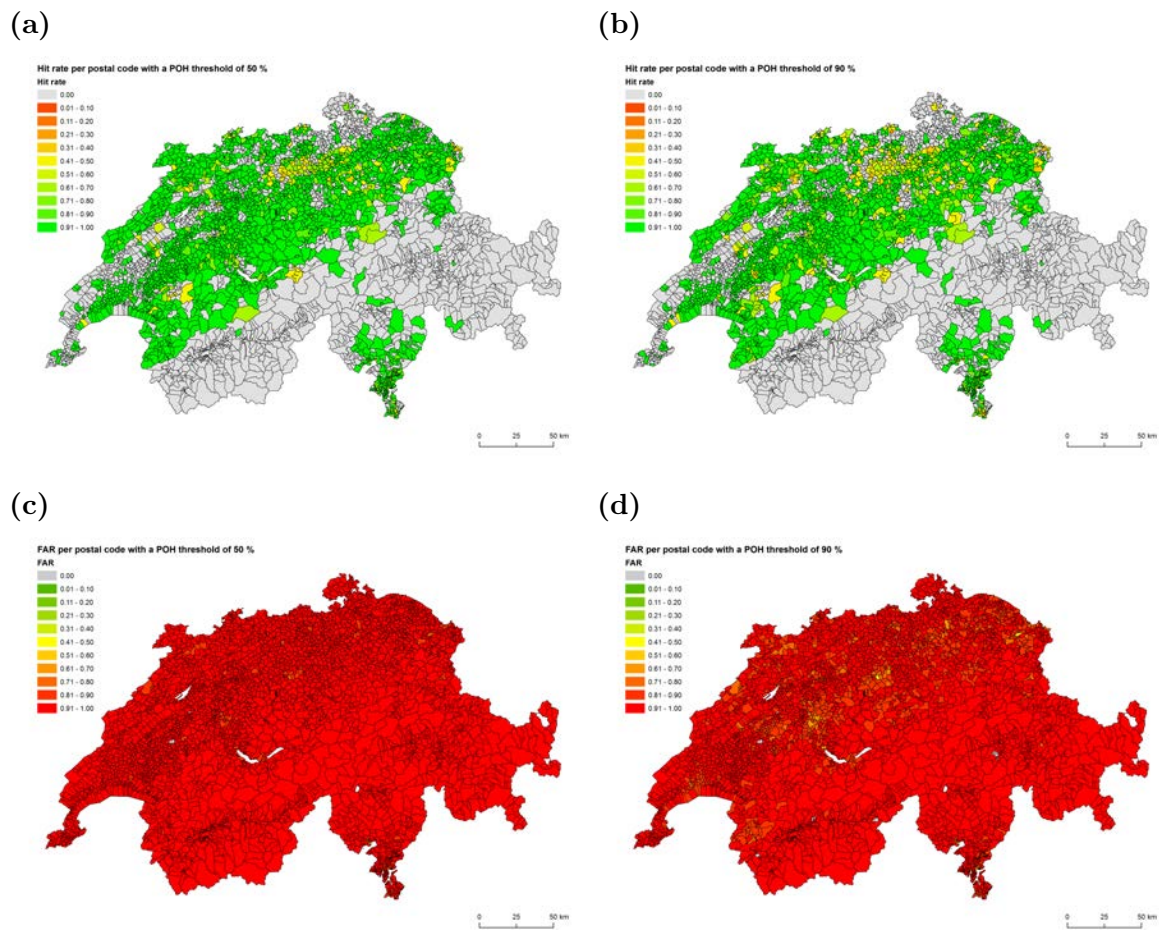
## Results per postal code areas

The Hit rate and the FAR are calculated for the POH thresholds of 50 and 90 % (Fig. 7). These thresholds have been selected as 50 % is representative for low POH values still leading to damage, while 90 % is representative for the highest POH values where most of the claims occur. Many postal code areas have a Hit rate of 0 because there is not enough insurance data on the ground to have a *hit* case in the postal code area using only these two POH thresholds (Fig. 7a and 7b). A Hit rate of 1 is also often the consequence of a lack of insurance data, as only one *hit* case is sufficient to retrieve a Hit rate of 1. It is still possible to see that the Hit rate mainly ranges between about 0.4 and 1. The Hit rate decreases in several postal code areas with the increase of the POH threshold from 50 to 90 %. The FAR ranges mainly between 0.8 and 1 for the POH threshold of 50 % and between 0.7 and 1 for the POH threshold of 90 % (Fig. 7c and 7d). This is also mostly due to the lack of insurance data, as many *false alarm* cases are present but only few *hit* cases in each postal code area are present. The lowest FAR is generally found in the most populated areas, where more cars are expected to be found.

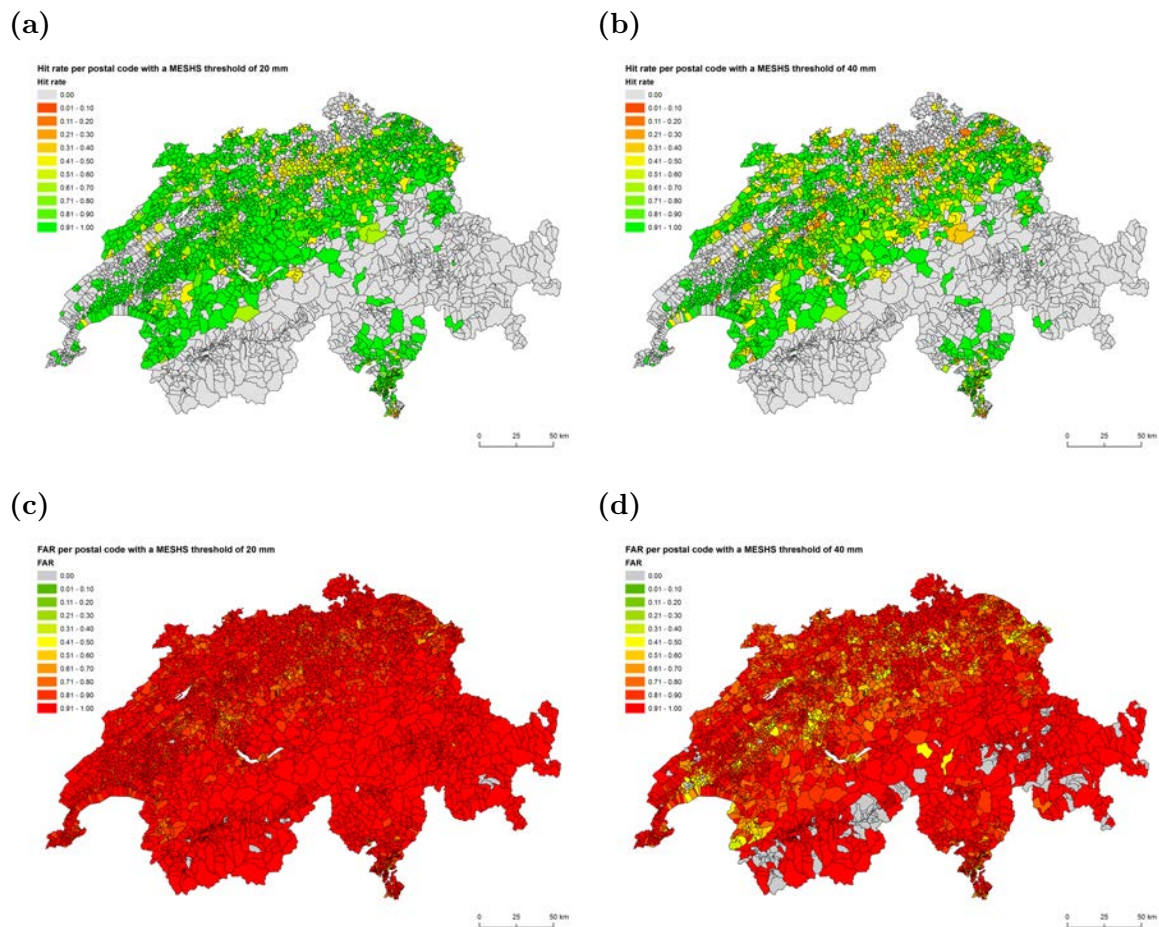
The Hit rate and the FAR are calculated for the MESHS thresholds of 20 and 40 mm (Fig. 8). These thresholds have been selected on the same basis as for the POH algorithm. Similarly to the POH algorithm many postal code areas have a Hit rate of 0 because there is not enough insurance data to have a *hit* case in the postal code area using only these two MESHS thresholds (Fig. 8a and 8b). A Hit rate of 1 is also often the consequence of a lack of insurance data, as only one *hit* case is sufficient to retrieve a Hit rate of 1. For a MESHS threshold of 20 mm, the Hit rate mainly ranges between 0.3 and 1. For a MESHS threshold of 40 mm, the range of the Hit rate is larger, corresponding to values between 0.1 and 1.

The FAR ranges mainly between 0.7 and 1 for the MESHS threshold of 20 mm (Fig. 8c). This range is larger for the MESHS threshold of 40 mm, with a FAR generally ranging between 0.4 and 1 (Fig. 8d). The general high FAR is also mostly due to the lack of insurance data, as many *false alarm* cases are present but only few *hit* cases in each postal code area. Similarly to the POH algorithm, the lowest FAR is generally found in the most populated areas, where more cars are expected to be found.

The results are more varied in the areas where more observations on the ground are expected. This is probably an effect of the lack of data in the other regions. It is still possible to see that the FAR slightly decreases in many postal code areas with the increase of the POH and MESHS thresholds. This is due to the fact that usually less radar pixels have a detection when taking into account only the highest POH/MESHS values. The same pattern is visible for the Hit rate but is not desired. In this case, the reason is the increase of *missed* cases with the reduction of radar detection.



**Figure 7:** Hit rate ((a) and (b)) and FAR ((c) and (d)) for the POH thresholds of 50 and 90 %. Results based on the dataset with insurance corrections using a threshold of 2 claims [graphics: A. Zischg].



**Figure 8:** Hit rate ((a) and (b)) and FAR ((c) and (d)) for the MESH thresholds of 20 and 40 mm. Results based on the dataset with insurance corrections using a threshold of 2 claims [graphics: A. Zischg].

# List of Figures

2.1	Daily maps for the POH (a) and MESHS (b) algorithms [MeteoSwiss]. . . . .	11
2.2	POH on the ground according to the difference of altitude between the height of the 45 dBZ and the freezing level [Waldvogel et al., 1979; Foote et al., 2005].	11
2.3	Nomogram of MESHS on the ground according to the difference of altitude between the height of the 50 dBZ and the freezing level [Treloar, 1998]. . . . .	12
3.1	Flow chart of the methodology. . . . .	14
3.2	POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 1 <sup>st</sup> of July 2012 [graphics: L. Nisi]. . .	16
3.3	2 × 2 contingency table for the verification of radar-based hail detection algorithms, based on the positive/negative detections and positive/negative observations. . . . .	17
3.4	POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 8 <sup>th</sup> of May 2003 [graphics: L. Nisi]. . .	20
3.5	POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 8 <sup>th</sup> of May 2003 (a) and on the 9 <sup>th</sup> of May 2003 (b) [graphics: L. Nisi]. . . . .	22
3.6	POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 25 <sup>th</sup> of May 2006 (a) and on the 25 <sup>th</sup> of June 2006 (b) [graphics: L. Nisi]. . . . .	23
3.7	POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 8 <sup>th</sup> of April 2003 [graphics: L. Nisi]. . .	23
3.8	POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 17 <sup>th</sup> of July 2010 (a) [graphics: L. Nisi] compared to the precipitation radar detection using daily maximum values (b) [MeteoSwiss]. . . . .	24
3.9	POH daily maximum value per pixel (color shading) and number of claims per postal code area (black squares) on the 1 <sup>st</sup> of August 2007 [graphics: L. Nisi]. .	24
3.10	Mask applied on the dataset and showing the percentage of traffic area per pixel for Switzerland [graphics: L. Thomi]. . . . .	26

---

4.1	Hit rate for different POH thresholds for the original dataset, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum (a), on the 85 <sup>th</sup> percentile (b), on the 75 <sup>th</sup> percentile (c) and on the median (d) within a postal code area. . . . .	30
4.2	FAR for different POH thresholds for the original dataset, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum (a), on the 85 <sup>th</sup> percentile (b), on the 75 <sup>th</sup> percentile (c) and on the median (d) within a postal code area. . . . .	31
4.3	Hit rate and FAR for different POH thresholds for the original dataset, using a threshold of 2 claims and based on the POH maximum within a postal code area. . . . .	31
4.4	Hit rate for different MESHS thresholds for the original dataset, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHS maximum (a), on the 85 <sup>th</sup> percentile (b), on the 75 <sup>th</sup> percentile (c) and on the median (d) within a postal code area. . . . .	32
4.5	FAR for different MESHS values for the original dataset, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHS maximum (a), on the 85 <sup>th</sup> percentile (b), on the 75 <sup>th</sup> percentile (c) and on the median (d) within a postal code area. . . . .	33
4.6	Number of claims for each day of the month for the time period 2003-2012, based on the dataset with insurance corrections. . . . .	34
4.7	Hit rate for different POH thresholds for the dataset with insurance corrections, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum values (a) and on the median (b) within a postal code area. . . . .	36
4.8	Hit rate for different POH thresholds for the original dataset and the dataset with insurance corrections, taking all the claims into account (a) or using a threshold of 2 claims (b). Results based on the POH maximum values within a postal code area. . . . .	36
4.9	FAR for different POH thresholds for the original dataset and the dataset with insurance corrections, taking all the claims into account (a) or using a threshold of 2 claims (b). Results based on the POH maximum values within a postal code area. . . . .	36
4.10	FAR for different POH thresholds for the dataset with insurance corrections, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum values (a) and on the median (b) within a postal code area. . . . .	37
4.11	Hit rate for different MESHS thresholds for the dataset with insurance corrections, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHS maximum values (a) and on the median (b) within a postal code area. . . . .	37



4.12 Hit rate for different MESHS thresholds for the original dataset and the dataset with insurance corrections, taking all the claims into account (a) or using a threshold of 2 claims (b). Results based on the POH maximum values within a postal code area. . . . . 37

4.13 FAR for different MESHS thresholds for the dataset with insurance corrections, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHS maximum values (a) and on the median (b) within a postal code area. . . . . 38

4.14 FAR for different MESHS thresholds for the original dataset and the dataset with insurance corrections, using a threshold of 2 claims, based on the MESHS maximum values (a) and on the 50<sup>th</sup> percentiles of MESHS values (b). . . . . 38

4.15 Hit rate for different POH thresholds for the original dataset, the dataset with insurance corrections and datasets with the masks, using a threshold of 2 claims. Results based on the POH maximum values within a postal code area. 39

4.16 Histograms showing the frequency of claims for each of the POH classes, using the dataset with insurance corrections. The frequency is shown taking all the claims into account (a), using a threshold of 2 and 5 claims ((b) and (c)). . . . 44

4.17 Total number of claims for each POH class for the period 2003-2012, based on the dataset with insurance corrections. . . . . 45

4.18 Histograms showing the frequency of claims for each of the MESHS classes, using the dataset with insurance corrections. The frequency is shown taking account all the claims into (a), using a threshold of 2 and 5 claims ((b) and (c)). 45

4.19 Total number of claims for each MESHS class for the period 2003-2012, based on the dataset with insurance corrections. . . . . 46

5.1 Comparison of the Hit rate (a) and of the FAR (b) of the POH algorithm for the POH maximum, the 85<sup>th</sup>, the 75<sup>th</sup> and the 50<sup>th</sup> percentiles of POH values for the original dataset using a threshold of 2 claims. . . . . 49

5.2 Masks applied on the dataset with insurance corrections and showing the pixels with at least 1 % (a), 5 % (b) and 50 % (c) of traffic area per pixel for Switzerland [graphics: L. Thomi]. . . . . 51

1 Hit rate for different POH thresholds for the dataset with the masks 1 %, 5 % and 50 %, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum ((a), (c) and (e)) and on the median ((b), (d) and (f)) within a postal code area. . . . . 62

2 FAR for different POH thresholds for the dataset with the masks 1 %, 5 % and 50 %, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum ((a), (c) and (e)) and on the median ((b), (d) and (f)) within a postal code area. . . . . 63

3	Hit rate for different MESHs thresholds for the dataset with the masks 1 %, 5 % and 50 %, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHs maximum ((a), (c) and (e)) and on the median ((b), (d) and (f)) within a postal code area. . . . .	64
4	FAR for different MESHs thresholds for the dataset with the masks 1 %, 5 % and 50 %, with a distinction for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHs maximum ((a), (c) and (e)) and on the median ((b), (d) and (f)) within a postal code area. . . . .	65
5	FAR for different POH thresholds for the dataset with insurance corrections “removed postal codes without claims” ((a) and (c)) and for the dataset with the mask 1 % “removed postal codes without claims” ((b) and (d)). A distinction is made for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the POH maximum ((a) and (b)) and on the median ((d) and (c)) within a postal code area. . . . .	66
6	FAR for different MESHs thresholds for the dataset with insurance corrections “removed postal codes without claims” ((a) and (c)) and for the dataset with the mask 1 % “removed postal codes without claims” ((b) and (d)). A distinction is made for the different thresholds of claims (at least 1, 2 or 5 claims). Results based on the MESHs maximum ((a) and (b)) and on the median ((c) and (d)) within a postal code area. . . . .	67
7	Hit rate ((a) and (b)) and FAR ((c) and (d)) for the POH thresholds of 50 and 90 %. Results based on the dataset with insurance corrections using a threshold of 2 claims [graphics: A. Zischg]. . . . .	69
8	Hit rate ((a) and (b)) and FAR ((c) and (d)) for the MESHs thresholds of 20 and 40 mm. Results based on the dataset with insurance corrections using a threshold of 2 claims [graphics: A. Zischg]. . . . .	70

# List of Tables

4.1	Hit rate of the POH algorithm for the different datasets. Results based on the POH maximum values, using a threshold of 2 claims. . . . .	42
4.2	Hit rate of the POH algorithm for the different datasets. Results based on the MESHS maximum values, using a threshold of 2 claims. . . . .	43



# Bibliography

- Austin, P. M. (1987). Relation between measured radar reflectivity and surface rainfall. *Monthly Weather Review*, 115:1053–1070.
- Berthet, C., Wesolek, E., Dessens, J., and et al. (2012). Extreme hail day climatology in Southwestern France. *Atmospheric Research*.
- Betschart, M. and Hering, A. (2012). Automatic Hail Detection at MeteoSwiss - Verification of the radar-based hail detection algorithms POH, MESHS and HAIL. Technical Report 238, Bundesamt für Meteorologie und Klimatologie, MeteoSchweiz.
- Brimelow, J. C. and Reuter, G. W. (2006). Spatial Forecasts of Maximum Hail Size Using Prognostic Model Soundings and HAILCAST. *Weather and Forecasting*, 21:206–219.
- Delobbe, L., Hollemand, I., Dehenauw, D., and Neméghaire, J. (2005). Verification of radar-based hail detection product. In *Preprints W.W.R.P. Symposium on Nowcasting and Very Short Range Forecasting (WSN05)*, Toulouse, France. P8.07.
- Delrieu, G., Creutin, J.-D., and Saint-Andre, I. (1991). Mean K-R relationships: Practical results for typical weather radar wavelengths. *J. Atmos. Oceanic Technol.*, 20:467–476.
- Foote, G. B., Krauss, T. W., and Makitov, V. (2005). Hail metrics using conventional radar. In *16<sup>th</sup> Conference on Planned and Inadvertent Weather Modification*, pages 1–6, San Diego, California. American Meteorological Society: Boston, MA. 10-13 January.
- Germann, U., Galli, G., Boscacci, M., and Bolliger, M. (2006). Radar precipitation measurement in a mountainous region. *Q. J. R. Meteorol. Soc.*, 132:1669–1692.
- Hering, A. M., Morel, C., Galli, G., Sényesi, S., Ambrosetti, P., and Boscacci, M. (2004). Nowcasting thunderstorms in the Alpine region using a radar based adaptive thresholding scheme. In *ERAD 2004*.
- Holleman, I., Wessels, H. R. A., Onvlee, J. R. A., and Barlag, S. J. M. (2000). Development of a hail-detection-product. *Phys. Chem. Earth (B)*, 25:1293–1297.
- Huntrieser, H., Schiesser, H. H., Schmid, W., and Waldvogel, A. (1997). Comparison of Traditional and Newly Developed Thunderstorm Indices for Switzerland. *American Meteorological Society - Weather and Forecasting*, 12:108–125. Institute of Atmospheric Science, Swiss Federal Institute of Technology, Zurich, Switzerland.

- Joss, J., Schädler, B., Galli, G., Cavalli, R., Boscacci, M., Held, E., Della Bruna, G., Kappenberger, G., Nespor, V., and Spiess, R. (1997). Operational Use of Radar for Precipitation Measurements in Switzerland. MeteoSvizzera, Locarno Monti, 1999.
- Joss, J. and Waldvogel, A. (1990). Precipitation measurement and hydrology. In *Radar in Meteorology: Battan Memorial and 40<sup>th</sup> Anniversary Radar Meteorology Conference*, page 577–597, Ed. D. Atlas. American Meteorological Society.
- Kunz, M. and Kugel, P. I. S. (2014). Detection of Hail Signatures from Single-polarization C-Band Radar Reflectivity. To be published in 2014.
- Kunz, M. and Puskeiler, M. (2010). High-resolution assessment of the hail hazard over complex terrain from radar and insurance data. *Meteorologische Zeitschrift*, 19(5):427–439. Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology (KIT), Germany.
- Kunz, M., Sander, J., and Kottmeier, C. (2009). Recent trends of thunderstorm and hail-storm frequency and their relation to atmospheric characteristics in southwest Germany. *International Journal of Climatology*, (29):2283–2297.
- Mahoney, K., Alexander, M. A., Thompson, G., Barsugli, J. J., and Scott, J. D. (2012). Changes in hail and flood risk in high-resolution simulations over Colorado’s mountains. *Nature climate change*, 2:125–131.
- Merino, A., García-Ortega, E., López, L., and et al. (2013). Synoptic environment, mesoscale configurations and forecast parameters for hailstorms in Southwestern Europe. *Atmospheric Research*, (122):183–198.
- Mohr, S. and Kunz, M. (2012). Recent trends and variabilities of convective parameters relevant for hail events in Germany and Europe. *Atmospheric Research*.
- Nisi, L. (2013). Hailstorms characteristics and their convective environments in the Swiss Alpine region. Technical report, Oeschger Centre for Climate Change Research, Institute of Geography University of Bern and MeteoSwiss Locarno-Monti.
- Pellarin, T., Delrieu, G., Saulnier, G.-M., Andrieu, H., Vignal, B., and Creutin, J.-D. (2002). Hydrologic visibility of weather radar systems operating in mountainous regions: Case study for the Ardèche catchment (France). *J. Hydrometeorol.*, 3:539–555.
- Schuster, S. S., Blong, R. J., and McAneney, K. J. (2006). Relationship between radar-derived hail kinetic energy and damage to insured buildings for severe hailstorms in Eastern Australia. *Atmospheric Research*, (81):215–235.
- Skripniková (2013). Relationship between radar-derived hail kinetic energy and damage to insured buildings for severe hailstorms in Eastern Australia. *Atmospheric Research*, (81):215–235.
- Smart, J. R. and Alberty, R. L. (1985). The NEXRAD hail algorithm applied to Colorado thunderstorms. In *Preprints, 14<sup>th</sup> Conf. on Severe Local Storms*, pages 244–247, Indianapolis, IN, Amer. Meteor. Soc.

- Studer, J. (2013). Überschwemmungen in der Schweiz seit 1984, Analyse von Versicherung Schadendaten. Master's thesis, Oeschger-Zentrum für Klimaforschung und Klimafolgenforschung, Geographisches Institut, Universität Bern.
- Treloar, A. (1998). Vertically integrated radar reflectivity as an indicator of hail size in the Greater Sydney region of Australia. In *19<sup>th</sup> Conference on Severe Local Storms*, pages 48–51, Minneapolis, Minnesota. 14-18 September.
- Waldvogel, A., Federer, B., and Grimm, P. (1979). Criteria for the detection of hail cells. *J. Appl. Meteor.*, (18):1521–1525.
- Wilks, D. S. (2006). *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series. Academic press, second edition.
- Witt, A., Eilts, M. D., Stumpf, G. J., Johnson, J. T., Mitchell, E. D. W., and Thomas, K. W. (1998). An enhanced hail detection algorithm for the WSR-88D. *Weat. Forecasting*, 13(2):286–303.
- Yuter, S. E. (2002). Precipitation radar. *Encyclopedia of Atmospheric Sciences*, page 833–1852.





# Acknowledgement

I particularly thank my supervisor Olivia Romppainen-Martius and my advisor Luca Nisi for their help, advice and support during this Master's Thesis. They were always available to answer my questions and guide me in my work as well as to support me when difficulties arose. I also thank the people from the Mobilab and more particularly Matthias Künzler, Luzius Thomi and Andreas Zischg for their support with the insurance data and for their ideas.

I also thank all the people from the fifth floor of the GIUB for the nice breaks and discussions which made the working days more pleasant. A big thank you as well to the IT people who helped me with my problems on the computer. Finally, I would like to thank my family and friends for their support and their proofreading of the Thesis.



# Declaration

under Art. 28 Para. 2 RSL 05

Last, first name: Morel, Sandrine

Matriculation number: 09-417-619

Programme: M.Sc. in Climate Sciences  
Bachelor  Master  Dissertation

Thesis title: Verification of radar-based hail detection algorithms with insurance loss data in Switzerland

Thesis supervisor: Prof. Dr. Olivia Romppainen-Martius

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due acknowledgement has been made in the text. In accordance with academic rules and ethical conduct, I have fully cited and referenced all material and results that are not original to this work. I am well aware of the fact that, on the basis of Article 36 Paragraph 1 Letter o of the University Law of 5 September 1996, the Senate is entitled to deny the title awarded on the basis of this work if proven otherwise.

Bern, August 25, 2014