https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

Modelling crop hail damage footprints with single-polarization radar: The roles of spatial resolution, hail intensity, and cropland density

Raphael Portmann[1], Timo Schmid[2,3], Leonie Villiger[2,3], David N. Bresch[2,3], and Pierluigi Calanca[1]

[1]Agroscope Reckhenholz, Climate and Agriculture, Zurich, Switzerland
[2]ETH Zurich, Institute for Environmental Decisions, Zurich, Switzerland
[3]Federal Office of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland

Correspondence: raphael.portmann@agroscope.admin.ch

Abstract. Hail remains a major threat to agriculture in Switzerland and beyond and assessments of current and future hail risk are of paramount importance for decision-making in the insurance industry and the agricultural sector. However, relating observational information on hail with crop-specific damages is challenging. Here, we build and systematically assess a model to predict hail damage footprints for field crops (wheat, maize, barley, rapeseed) and grapevine from the operational radar product Maximum Expected Severe Hail Size (MESHS) at different spatial resolutions. To this end, we combine the radar information with detailed geospatial information on agricultural land use and geo-referenced damage data from a crop insurer for 12 recent hail events in Switzerland. We find that for field crops, model skill gradually increases when the spatial resolution is reduced from 1 km down to 8 km. For even lower resolutions, the skill is diminished again. On the contrary, for grapevine, a lower model resolution tends to reduce skill, which is attributed to the different spatial distribution of field crops and grapevine in the landscape. It is shown that identifying a suitable MESHS thresholds to model damage footprints always involves trade-offs. For the lowest possible MESHS threshold (20 mm) the model predicts damage about two times too often (high frequency bias and number of false alarms) but also has a high probability of detection (80%). The frequency bias decreases for larger thresholds and reaches an optimal value close to 1 for MESHS thresholds of 30-40 mm. However, this comes at the cost of a substantially lower probability of detection (around 50%) while overall model skill remains largely unchanged. We argue that, ultimately, the best threshold selection therefore depends on the user need and the costs of a false alarm or a missed event. Finally, the frequency of false alarms can be substantially reduced when only areas with high cropland density are considered. Results from this simple, open-source model show that modelling of hail damage footprints to crops from single-polarization radar in Switzerland is skillful and is best done at 8 km resolution for field crops and 1 km for grapevine. They further allow different users of such models to identify the suitable threshold for their application, taking into account associated trade-offs.

1 Introduction

Hail storms frequently cause severe damage to agriculture and infrastructure in various places across the globe (Allen et al., 2020; SwissRe, 2022). In fact, severe convective storms (which include hailstorms) are among the costliest perils worldwide

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

(SwissRe, 2021). Larger-scale catastrophes like tropical cyclones or earthquakes cause substantially more damage per event, but severe convective (hail-)storms occur comparatively more often. Switzerland is a particularly hail prone country with an average of 32 hail days during the convective season (April to September) and locally up to three and more hail days in hot spot regions (Schroeer et al., 2022). The number of hail days varies strongly from year to year. Summer 2021 was an example for a record hail season causing extreme damages in a series of intense and widespread thunderstorms (Kopp et al., 2022). The main crop insurer in Switzerland, Schweizer Hagel (SH), reported around 14'000 damage claims and insured losses of around CHF 110 Million (approx. USD 117 million, Schweizer Hagel, 2021) for this year. Hail remains currently the costliest natural hazard for insured agricultural production in Switzerland and the events of summer 2021 demonstrated the need for reliable assessments of hail risk for key stakeholders such as, among others, insurers, governments, and farmers.

Here, we present a model to predict hail damage footprints to field crops and grapevine in Switzerland based on operational radar data and detailed information on agricultural land use. The model is verified with geo-referenced damage claims from SH. To make it accessible to stakeholders, the model is implemented in the open-source natural catastrophe modelling platform CLIMADA (Aznar-Siguan and Bresch, 2019).

Compared to other weather-related hazards, reliable data on hail remains scarce due to the small scale of thunderstorms and accompanying hail streaks as well as the high costs to maintain observational networks at a large scale. Therefore, radar data is frequently used to obtain an estimate of hail on the ground because it is continuous in space and time (Kunz and Kugel, 2015; Puskeiler et al., 2016). Recently, the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss) has compiled a comprehensive assessment of hail frequency and hail stone sizes for Switzerland based on 20 years of single-polarization radar data at a 1 km spatial resolution providing an important basis for hail risk assessments (Schroeer et al., 2022; Trefalt et al., 2022). The climatology is based on two hail products that are computed operationally: The Maximum Expected Severe Hail Size (MESHS) and the Probability of Hail (POH; Betschart and Hering, 2012). Both are based on a height difference between the melting level and the height of a certain reflectivity level, a criterion introduced by Waldvogel et al. (1979).

Single-polarization radar products such as MESHS and POH provide valuable estimates of the occurrence of hail on the ground as verified with, e.g. insurance claims (Holleman et al., 2000; Kunz and Kugel, 2015; Puskeiler et al., 2016; Nisi et al., 2016), reports of observers and media (Cică et al., 2015) and crowd-sourced hail reports (Barras et al., 2019). More specifically, verification based on insurance claims generally showed a high probability of detection [POD, fraction of damage events that is predicted, often around 0.8 or larger] but also relatively high false alarm ratios [FAR, fraction of predictions without damage, up to 0.8] (Kunz and Kugel, 2015; Puskeiler et al., 2016; Nisi et al., 2016; Warren et al., 2020; Schmid et al., 2023). However, these metrics usually strongly depend on the hail intensity threshold, the objective selection of which is not always possible. For example, one can choose the threshold with the highest skill score (e.g. Puskeiler et al., 2016) or one can require that frequency of damage prediction equals the frequency of damage occurrence (i.e. a frequency bias of 1, Warren et al., 2020). Further, these verification studies usually rely on pragmatic choices of the scale of spatial aggregation or the distance between damage claim and radar signal tolerated. Often, a reason for this was that insurance claims were only available at the community level, which does not allow assessing the skill at the km-scale. But a strong dependence of the skill on the spatial scale can be expected, as lowering the spatial resolution (or increasing the tolerated distance) yields a higher chance of overlap between radar signal and

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

damages (Holleman et al., 2000; Schmid et al., 2023). Also, information on the presence and density of cropland is essential for reliable skill metrics but has often not been incorporated in previous verification studies.

First efforts to relate crop damage of individual fields to radar information (Omoto and Seino, 1978; Seino, 1980; Schiesser, 1990) or to hail pad measurements (Changnon, 1971; Morgan, 1976; Katz and Garcia, 1981) derived crop-specific damage functions based on data pairs of damaged fields and observed measures of hail intensity. Schiesser (1990) presented damage functions that link harvest loss at the field scale for individual crop types at various phenological stages to hail kinetic energy derived from single-polarization radar. Directly applying such functions to radar measurements without without additional verification and taking into account cropland density will remain difficult to interpret. Sánchez et al. (1996) developed a statistical model to estimate harvest loss for barley and wheat based on hail sizes observed by hailpads and meteorological observers in northwestern Spain. More recent efforts used satellite imagery to estimate crop damage after a hail event (Bentley et al., 2002; Singh et al., 2017; Prabhakar et al., 2019; Bell et al., 2020; Sosa et al., 2021)

Despite these past efforts to quantify hail damages to specific crops, openly available models of this kind do currently, to our knowledge, not exist anywhere. Existing models have been developed in the insurance industry and are proprietary (AIR Worldwide, 2023). Hence, to build openly available (operational) radar-based hail damage models for individual crops, more systematic and open-source approaches are needed.

To extend the CLIMADA platform with a hail damage footprint detection module, in this study hail intensity measures from operational, single-polarization radar (MESHS and POH) are coupled to detailed, crop-specific, geo-referenced cropland information to build simple yes/no damage models for field crops (wheat, corn, barley, rapeseed) and grapevine at different spatial resolutions. The models are systematically verified based on detailed, crop-specific damage information from SH of twelve recent hail events in Switzerland.

More specifically, the following questions are addressed:

1. Which spatial resolution is most suitable to model hail damage footprints for field crops and grapevine based on operational, single-polarization radar data in Switzerland?

2. Is it possible to objectively define the best hail intensity threshold(s) to model hail damage footprints?

3. How sensitive is the model performance to cropland density?

While most of the study focuses on MESHS, the same methodology is also applied to POH and results are compared.

The remainder of this paper is structured as follows. First, the three main datasets are introduced, and the model setup and verification method described (Sect. 2). Then, it is discussed how the model skill depends on spatial resolution (Sect. 3.1) and MESHS threshold (Sect. 3.2). Subsequently, the combined effect of threshold and resolution is analyzed (Sect. 3.3). The sensitivity of this combined effect on cropland density and the selection of POH as hazard variable is then assessed in Sect. 3.4. Finally, the key results are discussed, including how the use of alternative verification approaches would affect them (Sect. 3.5). The paper ends with a summary of the key conclusions in Sect. 4.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

2 Data and Methods

2.1 Hail hazard data

In this study, single-polarization radar data products on a 1x1 km regular grid are used to characterize hail hazard. The Swiss radar network consists of 5 dual-polarization Doppler C-band radars and is in place in this form since 2016 (Germann et al., 2016). The two products used here, the Maximum expected severe hail size (MESHS), and the Probability of Hail (POH) are single-polarization radar products provided by MeteoSwiss (Betschart and Hering, 2012; Trefalt et al., 2022; Germann et al., 2022) and are computed operationally. MESHS depends on the empirical relationship between the size of the largest hailstone and the difference between the top of the 50 dBZ echo and the freezing level height. It is computed from the so-called "Treloar nomogram" of Joe et al. (2004), which is based on Treloar (1998). MESHS ranges from a minimum value of 20 mm and theoretically has no upper limit. Note that MESHS is designed to indicate the size of the largest hailstone within 1 km^2 and does not represent a spatial average. Although MESHS is not explicitly connected to actual hail size, positive relationships between crowd-sourced hail sizes and MESHS have been reported by Barras et al. (2019). Here, daily (06 UTC - 06 UTC) maximum MESHS values are used to define a hail day (or hail event). We use 06 UTC (08am local time) to define a hail day because it represents the minimum of the average daily hail activity (Schroeer et al., 2022). This minimizes the risk of splitting a hail storm into two consecutive hail days.

POH is based on an empirical relationship between the likelihood of hail at the ground and, similar to MESHS, the height difference between the top of the 45 dBZ echo and the environmental freezing level. It was originally introduced by Waldvogel et al. (1979) and further developed by Witt et al. (1998) and Foote et al. (2005). The form of the relationship by Foote et al. (2005) has been directly implemented by MeteoSwiss and is operational since 2008 (Trefalt et al., 2022). As for MESHS, daily (06 UTC - 06 UTC) maximum values are used.

To investigate how the model skill changes with reduced spatial resolution, the radar data (MESHS and POH) is aggregated to 2, 4, 8, 16, and 32 km regular grids using the maximum intensity value within each grid cell. The maximum is preferred over the mean because it largely conserves the value range of the intensity variable. Further, this approach is consistent with the assumption that the maximum intensity value determines the occurrence of damage.

2.2 Agricultural exposure data

Detailed geospatial information on agricultural land use was obtained from Swiss cantonal data on agricultural land use (geodienste.ch) that was then merged to a Swiss-wide dataset. The data used here are valid for 2021 and not available for earlier years but will be updated annually after 2021. The original data contain polygons of each agricultural field and information on its type of use and cultivated crop. For this study, the data are aggregated to the number of fields and the total crop area within a 1x1 km grid for winter wheat, maize (incl. silage and forage maize), winter barley, rapeseed, and grapevine, based on the centre point of each field. Further, an aggregate category called field crops is defined that incorporates winter wheat, maize, winter barley and rapeseed. To model yes/no hail damage footprints at the grid scale, the exposure is turned into a binary field

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

depending on the number of fields n within a grid cell

$$exposure = \begin{cases} 0 & n < n_{thresh} \\ 1 & n \geq n_{thresh} \end{cases} \quad (1)$$

If not specified otherwise, n_{thresh}, is set to 1. This means that a grid cell is considered as exposure if it contains at least one field of the considered crop type. Because it is expected that the probability of damage increases with n_{thresh}, the sensitivity of the model skill to different choices of n_{thresh} is examined in Sect. 3.4. Cropland density can be expressed in number of fields per km^2 (cropland number density, shown in Fig. A1) or, as also the area of each field is known, as the fraction of land area covered by a specific crop (cropland area fraction). For those grid cells containing at least one field at 1 km resolution the average cropland number densities are: 36.7 (grapevine), 9.7 (field crops), 4.6 (wheat), 4.3 (corn), 2.5 (barley), and 2.4 (rapeseed). The corresponding cropland area fractions are: 7.3% (grapevine), 14.4% (field crops), 5.7% (corn), 3.5% (barley), and 4.2% (rapeseed). The gridded cropland data (number of fields, total area) is provided open-source via the CLIMADA data API.

2.3 Model formulation

The model formulation evaluated here follows the risk framework of the IPCC (IPCC, 2022) implemented in CLIMADA (Aznar-Siguan and Bresch, 2019) and defines a hazard, exposure and a vulnerability (or impact function). The exposure consists of a binary field and the hazard consist of the radar data (here shown for MESHS). The impact function is defined by one threshold parameter s and represents a step function which is 0 below s and 1 above s.

$$f_{imp} = \begin{cases} 0 & MESHS < s \\ 1 & MESHS \geq s \end{cases} \quad (2)$$

The damage footprint then is computed as the product of exposure and impact function

$$damage_{footprint} = exposure \cdot f_{imp} \quad (3)$$

The resulting impact is a binary field of with 1 for grid cells where there is modeled damage and 0 where there is not.

2.4 Damage claims

To evaluate the skill of this model, damage information is obtained from damage claims provided by the Swiss Hail insurance company (SH) for ten hail days between 2017 and 2021. Damage information includes date, location, crop type and harvest loss as estimated by employees of SH in the field. Roughly a quarter of all claims indicated zero harvest loss and were removed. This resulted in a total of 26'292 crop-specific damage claims used for this study, out of which 21% are winter wheat, 26.5% maize (incl. grain and silage maize), 34% grapevine, 10% rapeseed, 8.5% winter barley. About 76 % of these claims contain explicit coordinates of the affected fields while the remaining claims are only provided at the community level. To still be

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

able to consider them in our analysis, these remaining claims are randomly distributed on all farmland (wheat, maize, barley, rapeseed) or all vineyards (grapevine) of that community. This procedure was repeated 1000 times for wheat and grapevine to assess the uncertainty associated to this random placement. It was found that the 95% confidence interval for the skill metrics considered in this study at 1 km spatial resolution is below 1% for wheat and below 2% for grapevine. Therefore, the uncertainty introduced by the random placement is considered small.

Based on a careful comparison with radar data and the Swiss Severe Weather Database (sturmarchiv.ch) some damage claims related to nocturnal hailstorms were re-dated to the previous day to match the time window of the radar data (06 UTC - 06 UTC). This resulted in a total of 12 hail days (see Table 1), instead of the 10 hail days provided in the original data by SH. Further, two hail events occurred when at least one of the crops had been completely or to a large extent harvested. These dates were identified if no or very few damage claims to a crop were registered, even if this crop type was present in a region where numerous damage claims are recorded for the other crops (see fourth column in Table 1). The obtained dates were further verified based on information on indicative starting dates for harvests (wheat: end of July, barley: end of June, rapeseed: mid-July, maize: October) (www.schweizerbauern.ch, 2023).

Finally, to allow damages to be compared directly to exposures and modelled damage footprints, damage claims were gridded to the same 1x1 km grid as the exposure data. This gridded dataset indicates the number of damaged fields separately for each crop type as well as the aggregate category field crops (wheat, maize, barley, rapeseed). It is important to note that average insurance coverage is 69% for field crops and 43% for grapevine (SH, personal communication) indicating that the total number of damaged fields is likely underestimated. This is expected to negatively affect model skill, mainly via a larger number of false alarms.

Damage and exposure data are from different sources and therefore it is checked whether damage actually occurs where exposure is identified. The fraction of claims that are in a 1 km grid cell without exposure is small (wheat: 3%, corn: 2%, barley: 7%, rapeseed 11%, field crops: 0.5%, grapevine: 0.4%) and reduces strongly for coarser resolutions. For the aggregate category field crops, the mismatch is significantly lower than for the individual crops. The more relevant number for our study is the fraction of grid cells with damage that have zero exposure, because such grid cells would reduce model skill. This fraction is larger but remains in the range of a few percent (wheat: 5%, corn: 4%, barley: 9%, rapeseed 13%, field crops: 1.5%, grapevine: 4%). This fraction also reduces for coarser resolutions, reaching almost zero at 8 km for all field crops and about 2% for grapevine. Hence, a coarser resolution can efficiently reduce the mismatch between damage and exposure, in particular for field crops. Note that the exposure used here is in principle only valid for 2021. However, there are only small differences in the mismatch between events in 2021 and events prior to 2021, indicating that this is not the major source of the mismatch. To avoid artificially reducing the skill of the model due to the (albeit small) mismatches, grid cells with damage but no exposure are excluded from the verification process. The gridded damage data (number of fields) is provided open-source via the CLIMADA data API.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

2.5 Verification based on contingency table

To measure the model skill at different spatial resolutions and for different hail intensity thresholds, a 2x2 contingency table is used (Table 2, c.f. Wilks, 2019). According to the model formulation, grid points with nonzero exposure that coincide with a hail intensity larger than a threshold s are considered damage predictions ($a+b$ in Table 2). Grid points with nonzero damages and nonzero exposure are considered damage observations ($a+c$ in Table 2). From the four numbers of the contingency table (a = hits, b = false alarms, c = misses, d = correct negatives) a range of scalar attributes and skill metrics are computed (Wilks, 2019):

$$\mathrm{FAR} = \frac{b}{a+b} \qquad \mathrm{POD} = \frac{a}{a+c}$$

$$\mathrm{B} = \frac{a+b}{a+c} = \frac{\mathrm{POD}}{1-\mathrm{FAR}} \qquad \mathrm{PC} = \frac{a+d}{a+b+c+d}$$

$$\mathrm{CSI} = \frac{a}{a+b+c} = \frac{1}{\frac{1}{1-\mathrm{FAR}} + \frac{1}{POD} - 1} \qquad \mathrm{HSS} = \frac{\mathrm{PC} - \mathrm{PC}_{rand}}{1-\mathrm{PC}_{rand}} = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)}$$

where FAR denotes the false alarm ratio (FAR=0 for a perfect model), POD the probability of detection or hit rate (POD=1 for a prefect model), B the frequency bias (perfect model: 1, overforecast: >1, underforecast: <1), PC the proportion correct, also called accuracy (PC=1 for a perfect model), CSI the critical success index, also known as threat score (CSI=1 for a perfect model), and HSS the Heidke Skill Score. PC and CSI are both measures of forecast accuracy, but CSI has the advantage that it is a simple measure to account for the trade-off between high POD and low FAR (Roebber, 2009). However, PC has the advantage that it takes into account the ability of the model to correctly predict non-events (correct negatives, d in Table 2). The HSS is a classical forecast skill score on the basis of PC and quantifies the accuracy of the forecast compared to a random forecast (perfect model: 1, no skill: 0, Heidke, 1926). The overall model performance in this study is assessed based on the HSS. For comparison, the radar product used by Warren et al. (2020) to compile a hail climatology for Brisbane and Sydney, Australia, achieved a HSS of around 0.5, which the authors regarded as 'moderate skill'.

3 Results

3.1 The effect of resolution on model skill

First, average model skill as measured by HSS across all events and its dependence on spatial resolution is investigated for each crop type individually for a MESHS threshold of s =20 mm (Fig. 1). For wheat, maize, rapeseed and barley model skill substantially increases up to 8 km and reduces or remains constant for higher resolutions. Aggregating them to one crop type (field crops) conserves this behaviour but increases overall skill. For grapevine, the behaviour is opposite: Skill reduces with coarser resolution up to 8 km and increases again thereafter. The increased skill when verifying hail damages on a larger scale is well known for neighbourhood-based approaches (Warren et al., 2020; Schwartz, 2017; Schmid et al., 2023), even if these approaches have substantial methodological differences to our resolution-based approach. It can essentially be explained by

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

the reduced penalization of forecasts due to e.g. spatial displacement of forecast and observation [hereafter referred to as scale effect]. However, the reduced skill with coarser resolution for grapevine cannot be explained with the scale effect.

To further explore this contrasting behaviour, the effect of a coarser resolution on model skill is considered for individual events for both wheat (Fig. 2) and grapevine (Fig. 3) (s =20 mm). Focusing on wheat at 1 km (Fig. 2a-c,d), high average POD and FAR (both around 0.8) are found, with considerable differences between events. We discuss three representative examples (a shifted forecast, an overforecast, and a good forecast) in more detail. An event with low skill (HSS=0.16, POD=0.47, FAR=0.89) occurred on 15 June 2019 over Western Switzerland (Fig. 2a). In this case, the observed damage footprint (grey and blue grid cells) is shifted to the east relative to the predicted damage footprint (red and blue grid cells), resulting in many misses and false alarms and few hits. Then, 28 June 2021 (Fig. 2b) was an extreme hail event with an exceptionally large spatial extent (Kopp et al., 2022) (HSS=0.21, POD=0.91, FAR=0.81). It is characterized by many false alarms, notably over North-Eastern Switzerland. This results in low skill, despite the high POD. With a large frequency bias (B=4.8) this forecast can be characterized as overforecast. Finally, the event on 12 July 2021 (Fig. 2c) has the best skill (HSS=0.51, POD=0.61, FAR=0.53). The main damage footprint over North-Eastern Switzerland was captured very well but a number of misses at the edges of the damage footprint and scattered over the Swiss Plateau lead to a lower POD compared to the previous examples.

Reducing the resolution to 8 km affects metrics of the three representative examples (Fig. 2d-g). The shifted forecast (15 June 2019, Fig. 2d) greatly improves (HSS=0.42) due to a substantially higher POD (0.93), a lower FAR (0.67). The overforecast (28 June 2021, Fig. 2e) also improves (HSS=0.38) but mostly because of a lower FAR (0.57) while POD remains unchanged. Finally, the skill of the good forecast (12 July 2021, Fig. 2f) remains largely unchanged, as the benefit of a lower FAR (0.38) is compensated by a lower POD (0.55). Again the individual scattered damage reports over the Swiss Plateau contribute substantially to this relatively low POD (without them, the POD would be 0.73). HSS increases for all 10 considered events if spatial resolution is reduced (Fig. 2g). However, there are substantial differences in the magnitude of the increase between events. The FAR reduces with lower resolution for all events and POD increases for 7 out of 10 events. POD does not increase for events where it is already very high (e.g., the overforecast) or where many misses are located far away from the modeled damage footprint (e.g. the good forecast).

For grapevine, the story is different (Fig. 3), as illustrated in the following with the 15 June 2019 event. At 1 km, the model predicts damage footprints for grapevine well (HSS=0.47; Fig. 3a,g). Reducing the resolution to 8 km in this case increases the FAR from 0.65 to 0.78 and reduces HSS to 0.30, despite a higher POD. A similar behaviour is observed for the 24 July 2021 event (Fig. 3c,f,g), while for the 28 June 2021 event, the skill remains unchanged (Fig. 3b,e,g). Considering all events the following overall pattern emerges (Fig. 3g): Reducing the resolution reduces HSS or does at least not increase it, despite a higher POD. This is in contrast to the behaviour for wheat. The difference mostly arises because FAR increases for grapevine but reduces for wheat. A key difference between the two considered crops is that the damage footprints for grapevine are more heterogeneous and scattered than for wheat due to the very localized distribution of grapevine in the landscape compared to the spatially more evenly spread distribution of wheat. Many false alarms appear in the regions with low density of grapevine while hits populate the regions with high grapevine density notably at the shores of the Lake Geneva, Neuchâtel and Biel in Western Switzerland and the Lake Zurich in the North-East. Reducing the resolution mainly increases the fraction of these

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

false alarms, leading to lower skill (Fig. 3d,f,g).

It is clear that the scale effect tends to reduce FAR with coarser resolution, irrespective of the crops spatial distribution. To understand the behaviour for grapevine, an additional effect of a coarser resolution on FAR has to be considered: The chance of a false alarm also depends on cropland density [herafter referred to as density effect]. The main reason for this is the enormous variability of hail within a storm at the scales of a few hundred meters (Morgan and Towery, 1975) combined with insured fractions of fields well below 100%. Hence, the average FAR at 1 km grid points with 1 field is higher (wheat: 82%, grapevine: 79%) than at gridpoints with 10 fields (wheat: 74%, grapevine: 60%). In conclusion, with coarser resolution, FAR is affected depending on how cropland density changes. For a crop that occurs everywhere in the domain, average cropland density within a grid cell is independent of the resolution (even if density within individual 1 km grid cells varies). However, the more a crop occurs fragmented in distinct parts of the domain the stronger cropland density decreases with coarser resolution. Hence, this density effect contributes to an increase of FAR. For wheat, the average cropland density within a grid cell decreases by slightly more than a factor of 2 from 4.6 fields per km^2 at 1 km to 1.9 fields per km^2 at 8 km resolution. For grapevine however, it decreases from 36.7 fields per km^2 to 4.2 fields per km^2 which is about a factor of 9 (note that the results are almost identical if cropland area fraction is used). The reason for these differences can also be expressed in terms of an areal inflation factor: Only every 9th 1 km grid cell within a 8 km grid cell actually contains exposure (inflation factor of 9). At all resolutions these inflation factors are much larger for grapevine than for wheat (Fig. A2).

In conclusion, the scale effect dominates over the density effect for wheat and the density effect dominates over the scale effect for grapevine. Hence, to achieve a good skill when modelling hail damage footprints it is beneficial to reduce the resolution from the original 1 km to about 8 km for field crops while 1 km provides the best skill for grapevine.

3.2 The effect of MESHS threshold on model skill

Next, we aim to identify suitable MESHS threshold(s) to model hail damage footprints for field crops and grapevine. An often used method to determine an ideal hail intensity threshold is to evaluate a skill metric (e.g. HSS or CSI) as function of threshold and determine the location of the maximum (see e.g. Puskeiler et al., 2016; Kunz and Kugel, 2015). However, for wheat and grapevine at 1 km resolution, CSI and HSS do not exhibit a clear maximum. They remain largely unchanged up to 40 mm for field crops and 35 mm for grapevine and decline at higher thresholds (Fig. 4). Note that the sample size for grapevine is substantially smaller, especially at large MESHS thresholds (grey bars in Fig. 4), leading to larger uncertainty of the exact skill values. Warren et al. (2020) suggest to additionally constrain the optimal threshold with the condition that B is close to 1 to avoid overforecasting. Here, this would result in an optimal threshold even above 40 mm for wheat and 45 mm for grapevine. A MESHS threshold of 30 mm for field crops would result in a frequency bias of 2, i.e. it results in twice as many forecasts than observations. Hence, selecting a threshold comes with a trade-off between (i) a high POD (blue line in Fig. 4) and (ii) a low FAR (red line in Fig. 4) and B closer to 1.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

3.3 Combined effects of resolution and threshold on model skill

To provide an overview of the combined effects of resolution and threshold on model skill, the performance diagram is used (Fig. 5; Roebber, 2009; Wilks, 2019). The performance diagram shows the relationship between POD and 1-FAR for model set-ups based on 1, 4 and 8 km spatial resolution and MESHS thresholds of 20, 30, and 40 mm. A perfect model is located in the top right of the diagram. For field crops it becomes evident that, for all three resolutions, an increase of the threshold strongly reduces POD but also reduces FAR and B (dashed diagonal lines), leaving its skill practically unchanged (shading, measured by CSI; Fig. 5a). Reducing the model resolution shifts the points in the diagram towards the top right, i.e. increases the skill by strongly reducing FAR, increasing POD and reducing B. Note that the more favourable skill measure, HSS, can not be shown in the performance diagram directly. However, CSI behaves similar to HSS for resolutions below 8 km.

The diagram also reveals the key differences between grapevine and field crops (Fig. 5b). Consistent with the results from Sect. 3.1, the main difference is that the FAR is not reduced with reduced resolution but even slightly increased, i.e. the "threshold-resolution web" is squeezed together in the horizontal direction. This results in a tendency for lower skill despite the small increase in POD.

3.4 Sensitivity to cropland density and hazard variable

Now, the sensitivities of the model performance to, first, cropland density (via n_{thresh}) and second, to the selection of an alternative hazard variable (POH) are discussed.

The sensitivity to cropland density is substantial. An increase in n_{thresh} leads to a decrease in FAR for all crops (Fig. 6), while POD remains largely unaffected (not shown). Until n_{thresh} values of about 20 for field crops at 8 km and 10 for grapevine at 1 km, the FAR decreases strongly by about 10 percent points for field crops and about 20 percent points for grapevine, respectively. Beyond these thresholds, the FAR-n_{thresh} curve tends to flatten out. Increasing n_{thresh} comes with the cost that the fraction of fields included is reduced (Fig. A3). Hence, an optimal value of n_{thresh} reduces FAR as much as possible but keeps the included fraction of fields or crop area high. Here, a pragmatic choice for all field crops at 8 km resolution is n_{thresh}=20, which maintains 95% of the number of fields for rapeseed, 96% for barley, 98% for wheat and maize and 99% for field crops (Fig. 6a). Note that, for certain crops, even higher n_{thresh} are justified (e.g. wheat, maize, field crops, see Fig. A3a). For the aggregate crop class field crops, even a n_{thresh}=100 (or larger) is justified, as it still preserves 96% of fields. Very similar numbers result when crop area is considered instead of field number. For grapevine at 1 km resolution, a suitable choice is n_{thresh}=10, which reduces FAR by almost 0.2 and still preserves 95% of the number of vineyards (Figs. 6b and A3b). However, it conserves only 86% of vineyard area, which would also justify a lower threshold.

The effect of a n_{thresh}=20 for field crops on model performance is illustrated in the performance diagram (Fig. 7a). For all resolutions, the "threshold-resolution-web" shifts to the right in the diagram compared to the original n_{thresh}=1. Hence, FAR is substantially reduced and POD remains nearly constant, leading to a higher skill. Note, however, that the choice of the optimal n$_t hresh$ heavily depends on the chosen spatial resolution.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

Finally, the sensitivity of the model performance to the selection of POH instead of MESHS, is tested (Fig. 7b). The identical analyses as for MESHS are performed for POH for thresholds of 70, 85, and 100% and 1, 4, and 8 km resolution and n_{thresh}=1. Compared to MESHS, the "threshold-resolution-web" is shifted towards the top left in the performance diagram. This indicates higher POD but also a higher FAR, and lower overall skill. These results are consistent with previous studies (Nisi et al., 2016; Schmid et al., 2023). Further, the highest possible threshold (100%) still exhibits substantial frequency biases (>1.5), limiting POH-based models to applications where overforcasting is not a problem. Hence, MESHS is better suited to model hail damage footprints for most applications.

3.5 Discussion

In this section, the key results are discussed in more detail, including how they would be affected by the use of different verification approaches.

The optimal resolution was found to differ for field crops (8 km) and grapevine (1 km). It was argued that two competing effects play a role: First, the scale effect tends to increase skill for coarser resolutions, because larger distances of forecast and damage observation are tolerated (Warren et al., 2020; Schwartz, 2017; Schmid et al., 2023). Second, the fact that the area covered by the exposure grid cells is artificially inflated with coarser resolution leads to lower cropland densities and hence, higher chances of false alarms, which reduces skill (density effect). The effect of a changing cropland density is particularly relevant, because hail storms are very localized phenomena with a high within-storm spatial variability (Morgan and Towery, 1975). The density effect strongly depends on the spatial distribution of crops: It is larger for crops that are scattered unevenly (like grapevine) and smaller for crops that occur more homogeneously distributed across the domain (like wheat and other field crops). Hence, reducing the spatial resolution is only beneficial for crops that are sufficiently evenly distributed in the landscape. We acknowledge that it remains open what ‘sufficiently’ is in this context. The dependence of cropland density on spatial resolution has also been acknowledged by (Griffith et al., 2000). In fact, it is a property that can be found for aggregation of any spatially heterogeneously distributed feature. For example, Baker et al. (2007) found that the density of drainage channels per unit area strongly reduced with coarser resolution.

Considering the MESHS threshold, the identified trade-off between high POD and a low FAR or a B close to 1 eventually signifies that the optimal threshold depends on user needs and the costs of a false alarm and a missed event. For example, if an insurance company wants to use this model to verify damage claims, it will prioritize a low threshold with a high POD. On the other hand, if scientists or governments use this model to communicate the damaged crop area after a hail event, they may want to avoid a systematic overestimation of the damage extent and chose a higher threshold. To incorporate the costs of false alarms and missed events in decision making with this model, user-tailored cost-loss models would have to be developed (de Elía, 2022). It is important to note that the best threshold for end users is not necessarily the one with the highest skill, but depends on their specific cost functions (Manzato, 2007).

The strongly reduced FAR with larger minimum number of fields within a grid cell (n_{thresh}) is again related to the large within-storm spatial variability of hail. The lower the cropland density, the higher the chances that a hail event does not lead to damage. Hence, hail damage footprints can be better modeled within the main crop production areas. These results are comparable to

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

Tian et al. (2018) who found that the FAR of satellite-based detection of rainfall occurrence reduces with increased rain gauge density.

Finally, it is noted that other, equally valid, verification procedures exist than the ones used in this study. Two alternatives and their effect on our results are briefly discussed. First, Ebert and Milne (2022) suggest the use of the Pierce Skill Score (PSS , Peirce, 1884) as alternative to HSS for rare and severe events. PSS favours a high POD and hence in our case, a MESHS threshold of 20 mm. Because of its high POD, a POH-based based model therefore outperforms a MESHS-based model when evaluated using PSS instead of HSS. PSS of the MESHS-based model for field crops remains nearly constant with coarser resolution until 8 km but reduces for even coarser resolutions, which corroborates the meaningfulness of the selection of an 8 km resolution.

Second, the use of fuzzy forecast verification has been proposed as alternative to point-based techniques to verify precipitation forecasts (Ebert, 2008). An often used fuzzy verification metric is the fractions skill score (FSS), which measures the fractional coverage of events in windows of different length scale around observations and forecasts (Roberts and Lean, 2008). It allows identifying the scale at which a forecast should be believed. Using a n_{thresh}=20, the 1, 2, and 4 km we find that models are skillful beyond a scale of 4 km and for MESHS thresholds between 20-30 mm. In general, the skill increases with larger scales and lower model resolutions. However, the 8 km model does not further increase FSS compared to the 4 km model. This perspective confirms that modeling hail damage footprints is not skillful at the 1 km scale but suggests that a 4 km resolution could also be a suitable choice. Considering grapevine (n_{thresh}=10), the lowest scale at which skillful prediction is possible is 6 km at a threshold of 20 mm. The FSS confirms that the skill does not improve with coarser spatial resolution, except for very large scales beyond 64 km.

4 Conclusions

This study presents an open-source model implemented in CLIMADA to predict the hail damage footprints (occurrence of hail damage yes/no) for individual crops after the passage of a hailstorm based on the operational single-polarization meteorological radar product MESHS and detailed agricultural land use data. Damage information from a crop insurer was used to quantify the skill of the model with different skill metrics. The main goal was to assess the model performance for different choices of spatial resolution (aggregation), MESHS threshold, and threshold of the number of fields used to define the exposure (cropland density).

For field crops (wheat, maize, rapeseed, barley) the model improves substantially when coarsening spatial resolution gradually from 1 km to 8 km, mainly because it relaxes the requirement for exact spatial overlap of modeled and observed damage footprints (scale effect). Beyond 8 km, the model skill tends to reduce again. On the contrary, for grapevine, coarser resolution tends to lower the model skill. We conclude that this difference between field crops and grapevine is mainly related to the different spatial distribution of these crops in the landscape [scattered for grapevine vs. more evenly for field crops] which determines how strongly cropland density reduces with coarser resolution. A lower cropland density leads to a higher chance of a false alarm (density effect). For wheat, the scale effect dominates, while for grapevine the density effect dominates.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

Increasing the hail intensity threshold from 20 mm to 40 mm strongly decreases the probability of damage detection (POD) but also reduces false alarm ratio (FAR) and frequency bias. The overall skill (HSS) remains largely unaffected by the threshold selection, due to the trade-off between POD and FAR that has to be aligned with user needs and their specific cost functions. Model performance can be substantially improved at all resolutions by selecting a higher minimum cropland density (n_{thresh}) for the exposure definition, mainly due to a reduction in FAR. Considering an alternative radar-based hail product (POH) reveals higher POD, higher FAR and lower skill compared to MESHS, confirming previous studies (Nisi et al., 2016; Schmid et al., 2023).

Finally, suitable model setups (resolution, threshold s, minimum exposure density n_{thresh}) can be identified to model damage footprints to field crops and grapevine. Some of the most suitable setups for field crops and grapevine are presented in Table 3 including associated skill metrics. For all crops, MESHS thresholds of 20 and 30 mm outperform MESHS thresholds of 40 mm in particular for higher n_{thresh}. The best performing setups for field crops are achieved at 8 km resolution and reach a POD of about 0.8 combined with a FAR of about 0.5 (for MESHS $>$ 20 mm) or a POD around 0.7 combined with an FAR about 0.4 (for MESHS $>$ 30 mm). For grapevine, the best performance is achieved at 1 km and reaches either POD of around 0.7 and a FAR of 0.6 (for MESHS $>$ 20 mm) or POD and FAR of around 0.55 (for MESHS $>$ 30 mm). These results are comparable to previous verification efforts of MESHS (Nisi et al., 2016) or the original Waldvogel et al. (1979) criterion (Puskeiler et al., 2016), although methodological verification approaches substantially differ from ours. Traditionally, verification of radar-based hail detection has focused on the dependence of the skill on the hail intensity threshold. Our work highlights that it is crucial to also consider the dependence on spatial scale and the density of cropland or, more generally, the observation network.

This simple damage yes/no model provides a first step towards the (operational) modelling of hail damages as well as hail risk assessments for crops in Switzerland. It is important to note that larger damage datasets would substantially increase the robustness of the results due to the large event-to-event variability. Gridded exposure and damage information are provided open-source via the CLIMADA data API to facilitate their use for operational purposes as well as the further development and validation of (hail) damage models for crops in Switzerland.

Code and data availability. The code (Python 3.9) to produce the figures in this manuscript and run the model is available at https://github.com/raphael-portmann/crop_hail_damage_footprint. Gridded exposure, damage and hazard information is available via the CLIMADA data API https://climada.ethz.ch/data-types/. CLIMADA is an open-source and -access software (https://doi.org/10.5281/zenodo.7691855) and can be used with any user-provided portfolio under the General Public Licence gpl-3.0.

Author contributions. All co-authors provided valuable comments and suggestions during in-depth discussions in the development of this study. Beyond that, specific contributions are: **RP**: Conceptualization of the study, methodology, software, visualizations, paper writing, review and editing. **TS**: methodology, software, paper review and editing. **LV**: paper review and editing. **PC**: conceptualization, paper review

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

Competing interests. The authors declare no competing interest.

Acknowledgements. The authors thank the Swiss hail insurance company (SH) for providing damage data for this study. In particular, we thank Guendalina Barloggio from SH for the helpful communication and discussions during the process. Further, we thank Hansueli Lusti and Hans Feyen from SH for valuable inputs and discussions. Finally, we thank colleagues from MeteoSwiss and the whole scClim team (https://scclim.ethz.ch/) for the many inputs and discussions. This study was funded by the Swiss National Science Foundation (SNSF) Sinergia grant CRSII5_201792.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

References

AIR Worldwide: AIR's Crop Hail Model, https://www.air-worldwide.com/models/crop/AIR-s-Crop-Hail-Model/, 2023.

Allen, J. T., Giammanco, I. M., Kumjian, M. R., Jurgen Punge, H., Zhang, Q., Groenemeijer, P., Kunz, M., and Ortega, K.: Understanding Hail in the Earth System, Rev. Geophys., 58, e2019RG000 665, https://doi.org/10.1029/2019RG000665, 2020.

Aznar-Siguan, G. and Bresch, D. N.: CLIMADA v1: A Global Weather and Climate Risk Assessment Platform, Geosci. Model. Dev., 12, 3085–3097, https://doi.org/10.5194/gmd-12-3085-2019, 2019.

Baker, M. E., Weller, D. E., and Jordan, T. E.: Effects of Stream Map Resolution on Measures of Riparian Buffer Distribution and Nutrient Retention Potential, Landscape Ecol, 22, 973–992, https://doi.org/10.1007/s10980-007-9080-z, 2007.

Barras, H., Hering, A., Martynov, A., Noti, P.-A., Germann, U., and Martius, O.: Experiences with >50,000 Crowdsourced Hail Reports in Switzerland, B. Am. Meteorol. Soc., 100, 1429–1440, https://doi.org/10.1175/BAMS-D-18-0090.1, 2019.

Bell, J. R., Gebremichael, E., Molthan, A. L., Schultz, L. A., Meyer, F. J., Hain, C. R., Shrestha, S., and Payne, K. C.: Complementing Optical Remote Sensing with Synthetic Aperture Radar Observations of Hail Damage Swaths to Agricultural Crops in the Central United States, J. Appl. Meteorol. Clim., 59, 665–685, https://doi.org/10.1175/JAMC-D-19-0124.1, 2020.

Bentley, M. L., Mote, T. L., and Thebpanya, P.: Using Landsat to Identify Thunderstorm Damage in Agricultural Regions, B. Am. Meteorol. Soc., 83, 363–376, https://doi.org/10.1175/1520-0477-83.3.363, 2002.

Betschart, M. and Hering, A.: Automatic Hail Detection at MeteoSwiss - Verification of the Radar- Based Hail Detection Algorithms POH, MESHS and HAIL, Arbeitsberichte der MeteoSchweiz, 238, 59, 2012.

Changnon, S. A.: Hailfall Characteristics Related to Crop Damage, J. Appl. Meteorol., 10, 270–274, 1971.

Cică, R., Burcea, S., and Bojariu, R.: Assessment of Severe Hailstorms and Hail Risk Using Weather Radar Data, Meteorol. Appl., 22, 746–753, https://doi.org/10.1002/met.1512, 2015.

de Elía, R.: The False Alarm/Surprise Trade-off in Weather Warnings Systems: An Expected Utility Theory Perspective, Environ. Syst. Decis., 42, 450–461, https://doi.org/10.1007/s10669-022-09863-1, 2022.

Ebert, E. E.: Fuzzy Verification of High-Resolution Gridded Forecasts: A Review and Proposed Framework, Meteorol. Appl., 15, 51–64, https://doi.org/10.1002/met.25, 2008.

Ebert, P. A. and Milne, P.: Methodological and Conceptual Challenges in Rare and Severe Event Forecast Verification, Nat. Hazard Earth Sys., 22, 539–557, https://doi.org/10.5194/nhess-22-539-2022, 2022.

Foote, B., Krauss, T., and Makitov, V.: Hail Metrics Using Conventional Radar, 85th AMS Annual Meeting, American Meteorological Society - Combined Preprints, 2005.

geodienste.ch: Land Use Map, https://geodienste.ch/services/lwb_nutzungsflaechen.

Germann, U., Boscacci, M., Gabella, M., and Schneebeli, M.: Weather Radar in Switzerland, in: From Weather Observations to Atmospheric and Climate Sciences in Switzerland, edited by Willemse, S. and Furger, M., VDF, https://doi.org/10.3929/ethz-a-010649833, 2016.

Germann, U., Boscacci, M., Clementi, L., Gabella, M., Hering, A., Sartori, M., Sideris, I. V., and Calpini, B.: Weather Radar in Complex Orography, Remote Sensing, 14, 503, https://doi.org/10.3390/rs14030503, 2022.

Griffith, J. A., Martinko, E. A., and Price, K. P.: Landscape Structure Analysis of Kansas at Three Scales, Landscape and Urban Plan., 52, 45–61, https://doi.org/10.1016/S0169-2046(00)00112-2, 2000.

Heidke, P.: Berechnung Des Erfolges Und Der Güte Der Windstärkevorhersagen Im Sturmwarnungsdienst, Geogr. Ann., 8, 301–349, https://doi.org/10.2307/519729, 1926.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

Holleman, I., Wessels, H. R. A., Onvlee, J. R. A., and Barlag, S. J. M.: Development of a Hail-Detection-Product: S10: Deep Convection, Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere, 25, 1293–1297, https://doi.org/10.1016/S1464-1909(00)00197-0, 2000.

IPCC: Climate Change 2022: Impacts, Adaptation and Vulnerability, Summary for Policymakers, Cambridge University Press, Cambridge, UK and New York, USA, 2022.

Joe, P., Burgess, D., Potts, R., Keenan, T., Stumpf, G., and Treloar, A.: The S2K Severe Weather Detection Algorithms and Their Performance, Weather Forecast., 19, 43–63, https://doi.org/10.1175/1520-0434(2004)019<0043:TSSWDA>2.0.CO;2, 2004.

Katz, R. W. and Garcia, R. R.: Statistical Relationships between Hailfall and Damage to Wheat, Agr. Meteorol., 24, 29–43, https://doi.org/10.1016/0002-1571(81)90031-5, 1981.

Kopp, J., Schröer, K., Schwierz, C., Hering, A., Germann, U., and Martius, O.: The Summer 2021 Switzerland Hailstorms: Weather Situation, Major Impacts and Unique Observational Data, Weather, https://doi.org/10.1002/wea.4306, 2022.

Kunz, M. and Kugel, P. I. S.: Detection of Hail Signatures from Single-Polarization C-band Radar Reflectivity, Atmos. Res., 153, 565–577, https://doi.org/10.1016/j.atmosres.2014.09.010, 2015.

Manzato, A.: A Note On the Maximum Peirce Skill Score, Weather Forecast., 22, 1148–1154, https://doi.org/10.1175/WAF1041.1, 2007.

Morgan, G. M.: Crop Damage-hailpad Parameter Study in Illinois, 1976.

Morgan, G. M. and Towery, N. G.: Small-Scale Variability of Hail and Its Significance for Hail Prevention Experiments, J. Appl. Meteorol. Clim., 14, 763–770, https://doi.org/10.1175/1520-0450(1975)014<0763:SSVOHA>2.0.CO;2, 1975.

Nisi, L., Martius, O., Hering, A., Kunz, M., and Germann, U.: Spatial and Temporal Distribution of Hailstorms in the Alpine Region: A Long-Term, High Resolution, Radar-Based Analysis, Q. J. Roy. Meteor. Soc., 142, 1590–1604, https://doi.org/10.1002/qj.2771, 2016.

Omoto, Y. and Seino, H.: On Relationships between Hailfall Characteristics and Crop Damage, J. Agric. Meteorol., 34, 65–76, https://doi.org/10.2480/agrmet.34.65, 1978.

Peirce, C. S.: The Numerical Measure of the Success of Predictions, Science, ns-4, 453–454, https://doi.org/10.1126/science.ns-4.93.453.b, 1884.

Prabhakar, M., Gopinath, K. A., Reddy, A. G. K., Thirupathi, M., and Rao, C. S.: Mapping Hailstorm Damaged Crop Area Using Multispectral Satellite Data, Egypt. J. Remote. Sens. Space Sci., 22, 73–79, https://doi.org/10.1016/j.ejrs.2018.09.001, 2019.

Puskeiler, M., Kunz, M., and Schmidberger, M.: Hail Statistics for Germany Derived from Single-Polarization Radar Data, Atmos. Res., 178–179, 459–470, https://doi.org/10.1016/j.atmosres.2016.04.014, 2016.

Roberts, N. M. and Lean, H. W.: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events, Mon. Weather Rev., 136, 78–97, https://doi.org/10.1175/2007MWR2123.1, 2008.

Roebber, P. J.: Visualizing Multiple Measures of Forecast Quality, Weather Forecast., 24, 601–608, https://doi.org/10.1175/2008WAF2222159.1, 2009.

Sánchez, J. L., Fraile, R., de la Madrid, J. L., de la Fuente, M. T., Rodríguez, P., and Castro, A.: Crop Damage: The Hail Size Factor, J. Appl. Meteorol., 35, 1535–1541, 1996.

Schiesser, H.: Hailfall: The Relationship between Radar Measurements and Crop Damage, Atmos. Res., 25, 559–582, https://doi.org/10.1016/0169-8095(90)90038-E, 1990.

Schmid, T., Portmann, R., Villiger, L., Schröer, K., and Bresch, D. N.: An Open-Source Radar-Based Hail Damage Model for Buildings and Cars, Nat. Hazard Earth Sys., pp. 1–38, https://doi.org/10.5194/nhess-2023-158, 2023.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

Schroeer, K., Trefalt, S., Hering, A., Germann, U., and Schwierz, C.: Hagelklima Schweiz: Daten, Ergebnisse Und Dokumentation, Tech. Rep. 283, MeteoSwiss, 2022.

Schwartz, C. S.: A Comparison of Methods Used to Populate Neighborhood-Based Contingency Tables for High-Resolution Forecast Verification, Weather Forecast., 32, 733–741, https://doi.org/10.1175/WAF-D-16-0187.1, 2017.

Schweizer Hagel: Extremwetterjahr 2021 – Ein Rückblick / Schweizer Hagel, https://www.hagel.ch/de/medien/extremwetterjahr-2021-ein-rueckblick/, 2021.

Seino, H.: On the Characteristics of Hail Size Distribution Related to Crop Damage, J. Agric. Meteorol., 36, 81–88, https://doi.org/10.2480/agrmet.36.81, 1980.

Singh, S. K., Saxena, R., Porwal, A., Ray, N., and Ray, S. S.: Assessment of Hailstorm Damage in Wheat Crop Using Remote Sensing, Curr. Sci., 112, 2095–2100, 2017.

Sosa, L., Justel, A., and Molina, Í.: Detection of Crop Hail Damage with a Machine Learning Algorithm Using Time Series of Remote Sensing Data, Agronomy, 11, 2078, https://doi.org/10.3390/agronomy11102078, 2021.

sturmarchiv.ch: Hagel – Schweizer Sturmarchiv, http://www.sturmarchiv.ch/index.php/Hagel#2010-2019.

SwissRe: Sigma 1/2021 - Natural Catastrophes in 2020, Tech. rep., SwissRe, 2021.

SwissRe: Severe 2022 Hail Damage in France Sets New Benchmarks, Underscores Shift of Risk and Calls for Pricing Adjustments, Tech. rep., SwissRe, 2022.

Tian, F., Hou, S., Yang, L., Hu, H., and Hou, A.: How Does the Evaluation of the GPM IMERG Rainfall Product Depend on Gauge Density and Rainfall Intensity?, J. Hydrometeorol., 19, 339–349, https://doi.org/10.1175/JHM-D-17-0161.1, 2018.

Trefalt, S., Germann, U., Hering, A., Clementi, L., Boscacci, M., Schröer, K., and Schwierz, C.: Hail Climate Switzerland Operational Radar Hail Detection Algorithms at MeteoSwiss: Quality Assessment and Improvement, Tech. Rep. 284, MeteoSwiss, 2022.

Treloar, A. B. A.: Vertically Integrated Radar Reflectivity as an Indicator of Hail Size in the Greater Sydney Region of Australia, in: Preprints, 19th Conf. on Severe Local Storms, Minneapolis, MN, Amer. Meteor. Soc, pp. 48–51, 1998.

Waldvogel, A., Federer, B., and Grimm, P.: Criteria for the Detection of Hail Cells, J. Appl. Meteorol., 18, 1521–1525, 1979.

Warren, R. A., Ramsay, H. A., Siems, S. T., Manton, M. J., Peter, J. R., Protat, A., and Pillalamarri, A.: Radar-Based Climatology of Damaging Hailstorms in Brisbane and Sydney, Australia, Q. J. Roy. Meteor. Soc., 146, 505–530, https://doi.org/10.1002/qj.3693, 2020.

Wilks, D. S.: Chapter 9 - Forecast Verification, in: Statistical Methods in the Atmospheric Sciences (Fourth Edition), edited by Wilks, D. S., pp. 369–483, Elsevier, https://doi.org/10.1016/B978-0-12-815823-4.00009-2, 2019.

Witt, A., Eilts, M. D., Stumpf, G. J., Johnson, J. T., Mitchell, E. D. W., and Thomas, K. W.: An Enhanced Hail Detection Algorithm for the WSR-88D, Weather Forecast., 13, 286–303, https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2, 1998.

www.schweizerbauern.ch: Ackerbau: Wissen & Facts, 2023.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

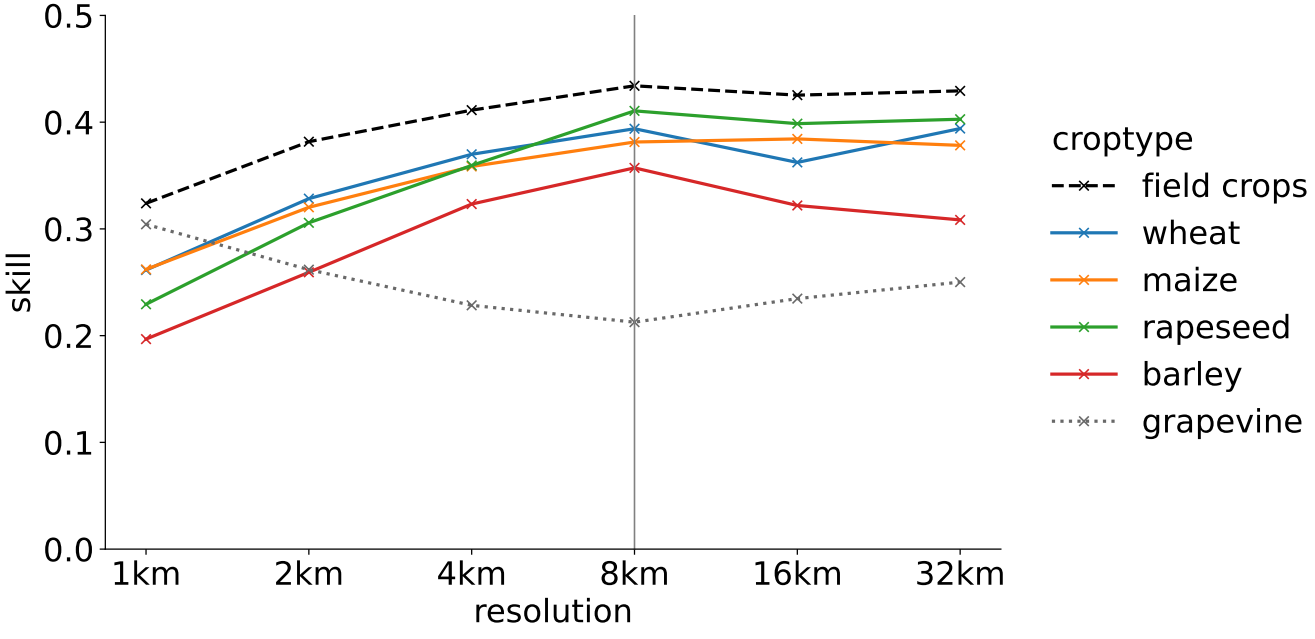

Figure 1. Skill (HSS) of the prediction of hail damage footprints with MESHS (20 mm threshold) for wheat, maize, rapeseed, barley and grapevine as well as an aggregate class field crops (includes wheat, maize, rapseed, barley) as a function of spatial resolution.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

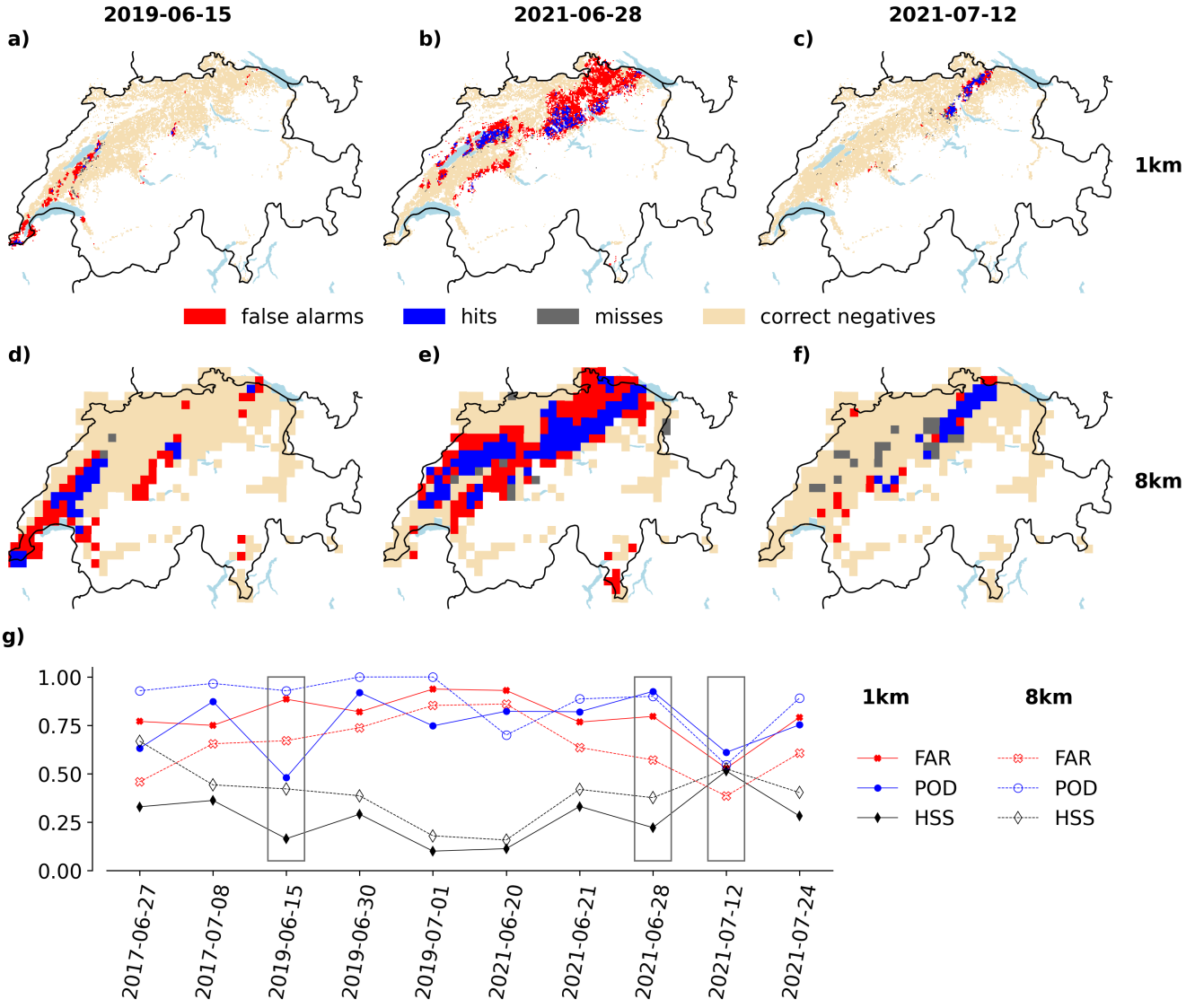

Figure 2. (a-c) 1x1 km and (d-f) 8x8 km grid cells classified as false alarms (red), hits (blue), and misses (grey) for damages to wheat based on MESHS $>$ 20 mm. Dates shown are (a,d) 15 June 2019, (b,e) 28 June 2021, and (c,f) 12 July 2021. (g) FAR, POD, HK, and HSS for all 10 recorded events and for (filled symbols) 1 km spatial resolution and (empty symbols) 8 km spatial resolution. Black boxes in panel (d) indicate the events shown in panels (a-f).

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

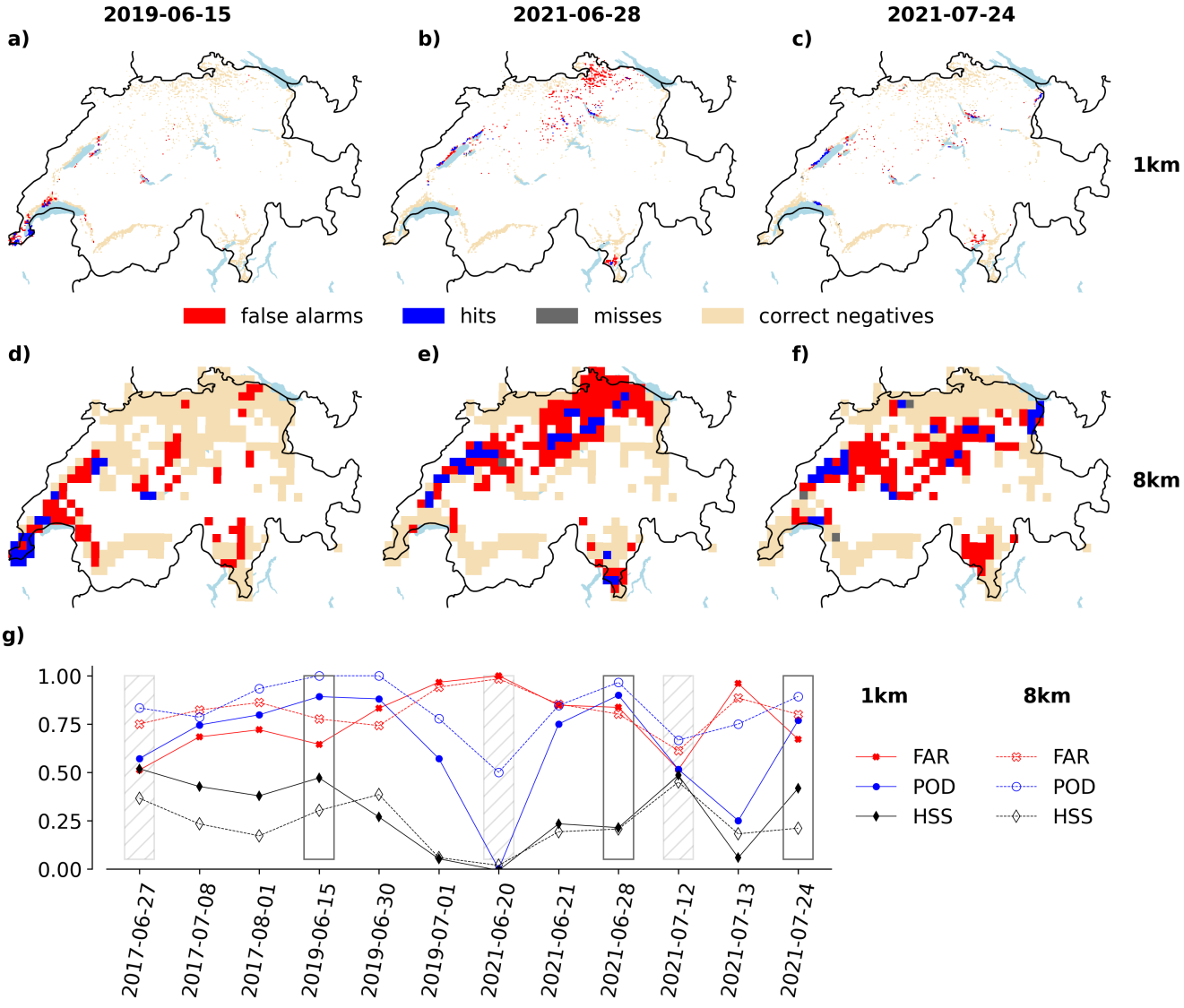

Figure 3. (a-c) 1x1 km and (d-f) 8x8 km grid cells classified as false alarms (red), hits (blue), and misses (grey) for damages to grapevine on (a) 15 June 2019, (b) 28 June 2021, and (c) 12 July 2021 based on MESHS $>$ 20 mm. (d) FAR, POD, HK, and HSS for all 12 recorded events and for (filled symbols) 1 km spatial resolution and (empty symbols) 8 km spatial resolution. Black boxes in panel (d) indicate the events shown in panels (a-f) and grey hatched boxes show events where the modeled damage footprint is smaller than 80km^2

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

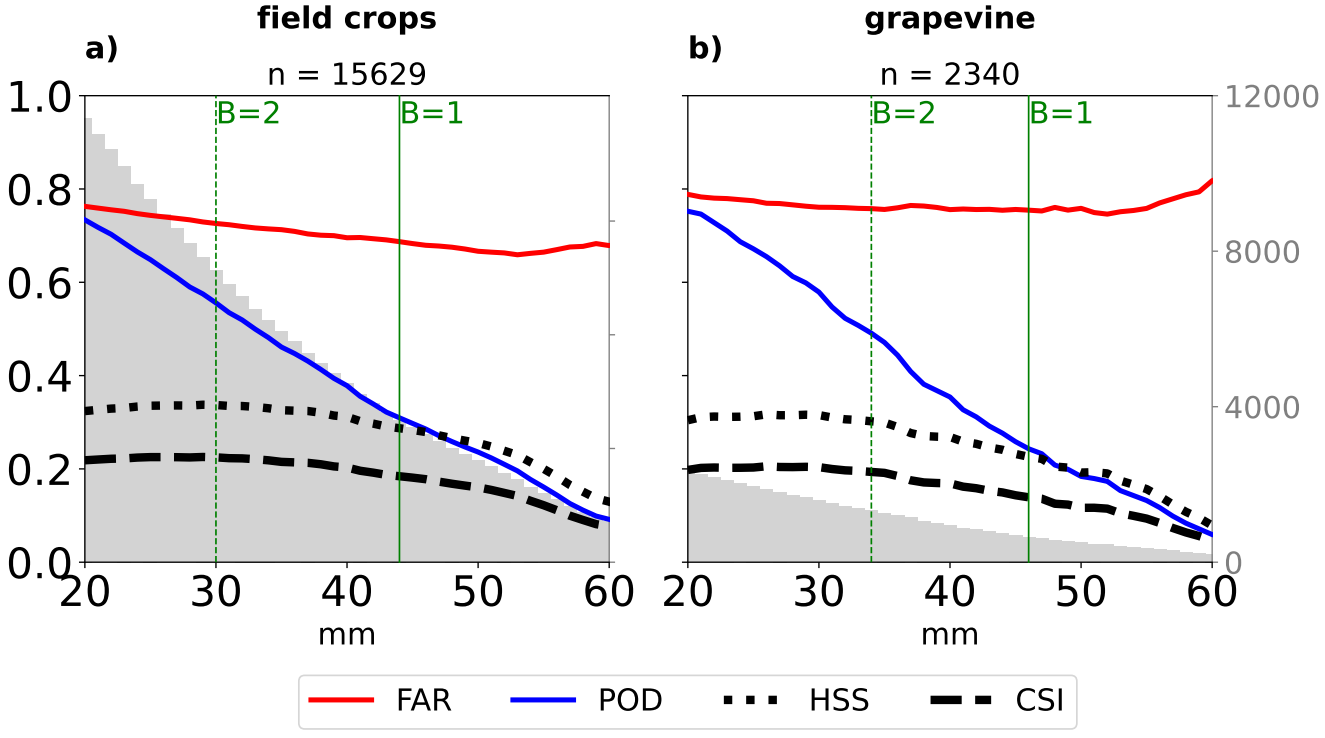

Figure 4. POD (blue), FAR (red), HSS (black dotted), CSI (black dashed) as a function of MESHS threshold for (a) field crops and (b) grapevine. Vertical green bars show thresholds with frequency biases (B) of approximately 1 and 2 and grey bars the total number of predictions at each threshold, as indicated by the vertical axis on the right. The total number of predictions summed over all thresholds is indicated at the top of each panel.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

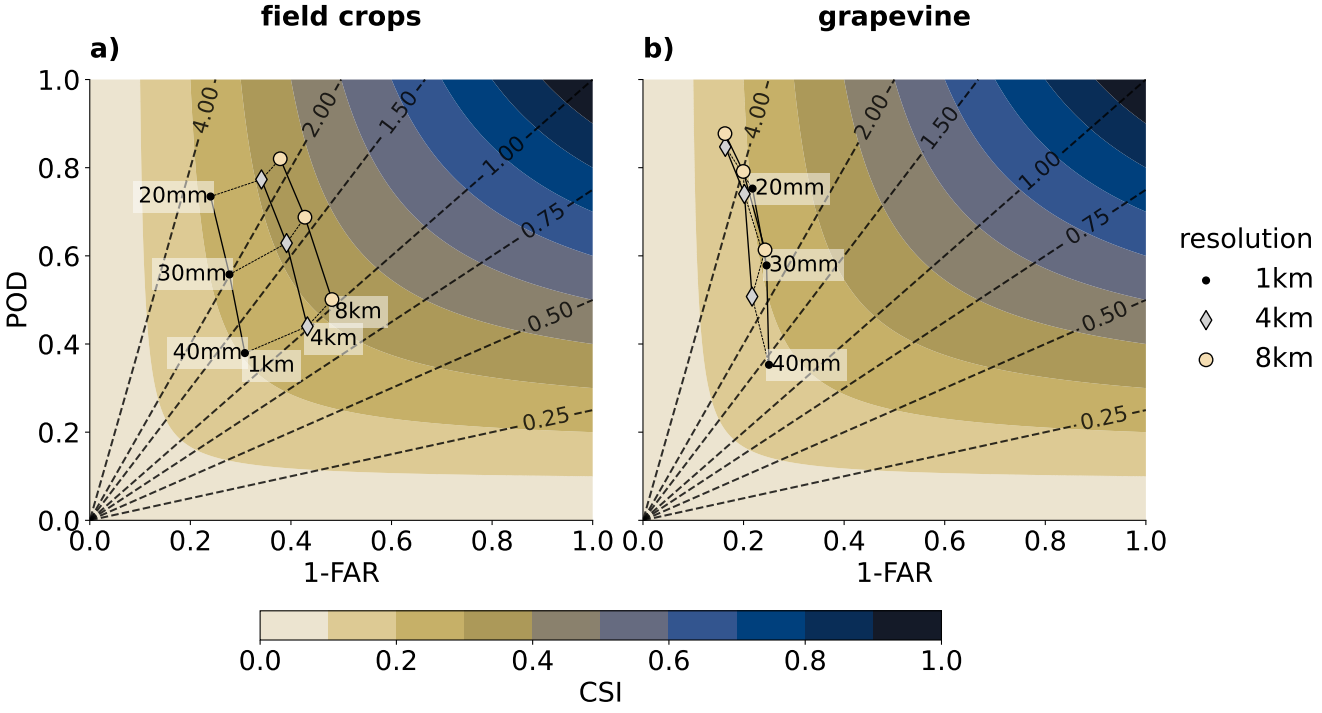

Figure 5. Performance diagrams showing POD vs. 1-FAR and CSI (shading) for (a) field crops and (b) grapevine for MESHS thresholds of 20, 30 and 40 mm and spatial resolutions of 1, 4, and 8 km. Dashed diagonal lines denote the frequency bias B.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

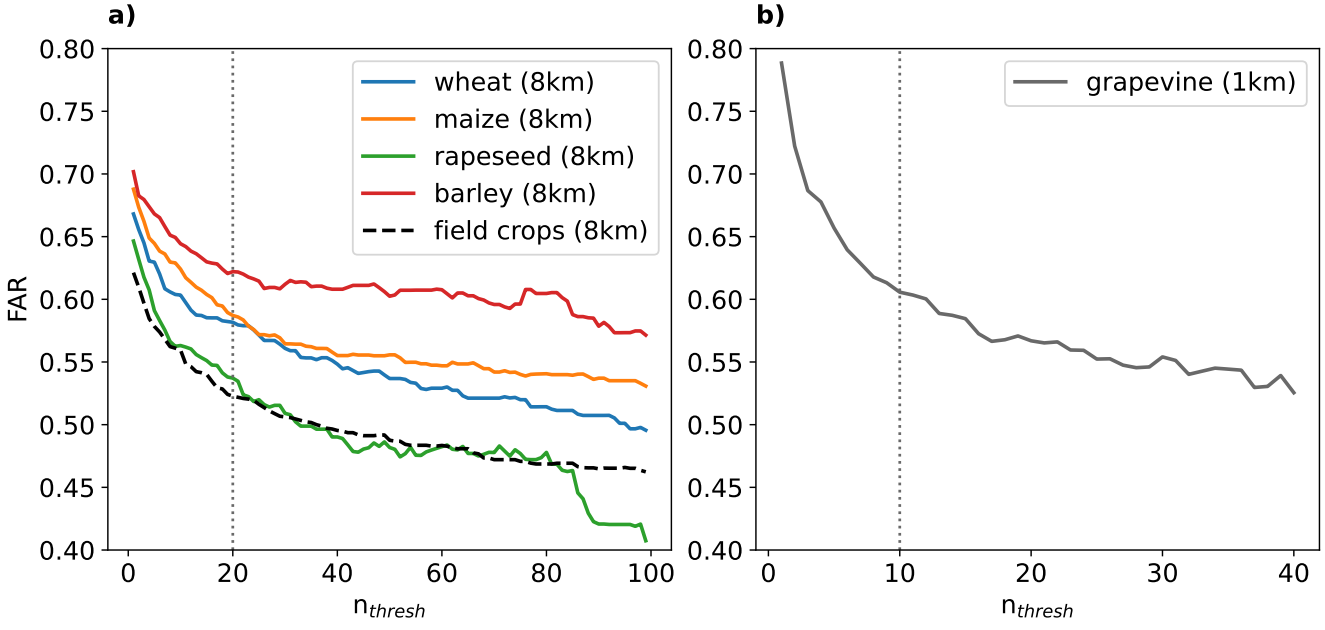

Figure 6. Change of the FAR as a function of n_{thresh} (number of fields per grid cell) for (a) wheat (blue), maize (orange), rapeseed (green), barley (red), and field crops (black, dashed) at 8 km resolution and for (b) grapevine at 1 km resolution. The vertical bars denote pragmatic choices of n_{thresh} that avoid too high FAR but still includes a large fractions (>95%) of total exposed crop area / number of fields.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

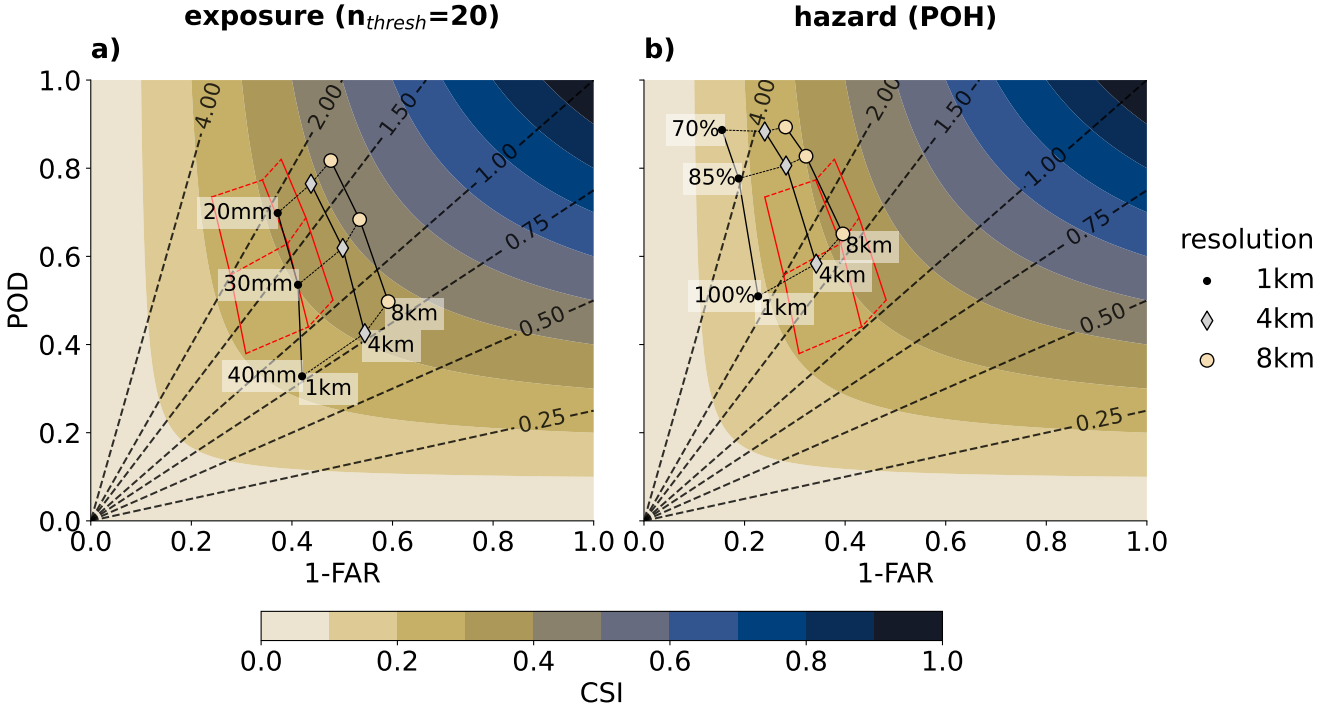

Figure 7. Performance diagrams for field crops for (a) alternative exposure, with an n_{thresh}=20 and (b) an alternative hazard (POH, 70, 85, and 100% thresholds). The values indicated with the red shape are shown for reference and are identical with the values shown in Figure 5a (MESHS, n_{thresh}=1).

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

Table 1. Overview of the 12 investigated hail days.

date	number of claims	damaged crops	harvested crops	comment
27 June 2017	2192	wheat, maize, barley, rapeseed, grapevine	-	
08 July 2017	2824	wheat, maize, barley, rapeseed, grapevine, apples	barley	
01 August 2017	1267	wheat, maize, rapeseed, grapevine, apples	wheat, barley, rapeseed	nocturnal hailstorm
15 June 2019	2185	wheat, maize, barley, rapeseed, grapevine, apples	-	
30 June 2019	632	wheat, maize, barley, rapeseed, grapevine	-	nocturnal hailstorm
01 July 2019	549	wheat, maize, barley, rapeseed, grapevine, apples	-	
20 June 2021	558	wheat, maize, barley, rapeseed, grapevine, apples	-	
21 June 2021	2228	wheat, maize, barley, rapeseed, grapevine, apples	-	
28 June 2021	7383	wheat, maize, barley, rapeseed, grapevine, apples	-	
12 July 2021	2109	wheat, maize, barley, rapeseed, grapevine, apples	-	nocturnal hailstorm
13 July 2021	96	maize, grapevine	-	weak hail event
24 July 2021	4269	wheat, maize, barley, rapeseed, grapevine, apples	-	

Table 2. 2x2 contingency table

		observations	
		yes	*no*
predictions	*yes*	a (hits)	b (false alarms)
	no	c (misses)	d (correct negatives)

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

Table 3. Most suitable model setups for field crops (aggregate crop class including wheat, maize, rapeseed, and barley) and grapevine.

crop type	parameters			skill metrics		
	resolution (km)	s (mm)	n_{thresh}	POD	FAR	HSS
field crops	8	20	100	0.80	0.46	0.55*
field crops	8	20	20	0.82	0.52	0.51
field crops	8	20	1	0.82	0.62	0.43
field crops	8	30	100	0.67	0.40	0.55*
field crops	8	30	20	0.68	0.47	0.52
field crops	8	30	1	0.68	0.57	0.45
field crops	8	40	100	0.47	0.34	0.47
field crops	8	40	20	0.50	0.41	0.47
field crops	8	40	1	0.50	0.52	0.43
grapevine	1	20	10	0.71	0.60	0.48*
grapevine	1	20	1	0.75	0.79	0.30
grapevine	1	30	10	0.55	0.56	0.46
grapevine	1	30	1	0.57	0.76	0.32

* highest skill for this crop type. Note that n_{thresh}=100 is only a sensible choice for the aggregate crop class field crops due to its high cropland density. For individual crop types, lower values like n_{thresh}=20 are to be preferred.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

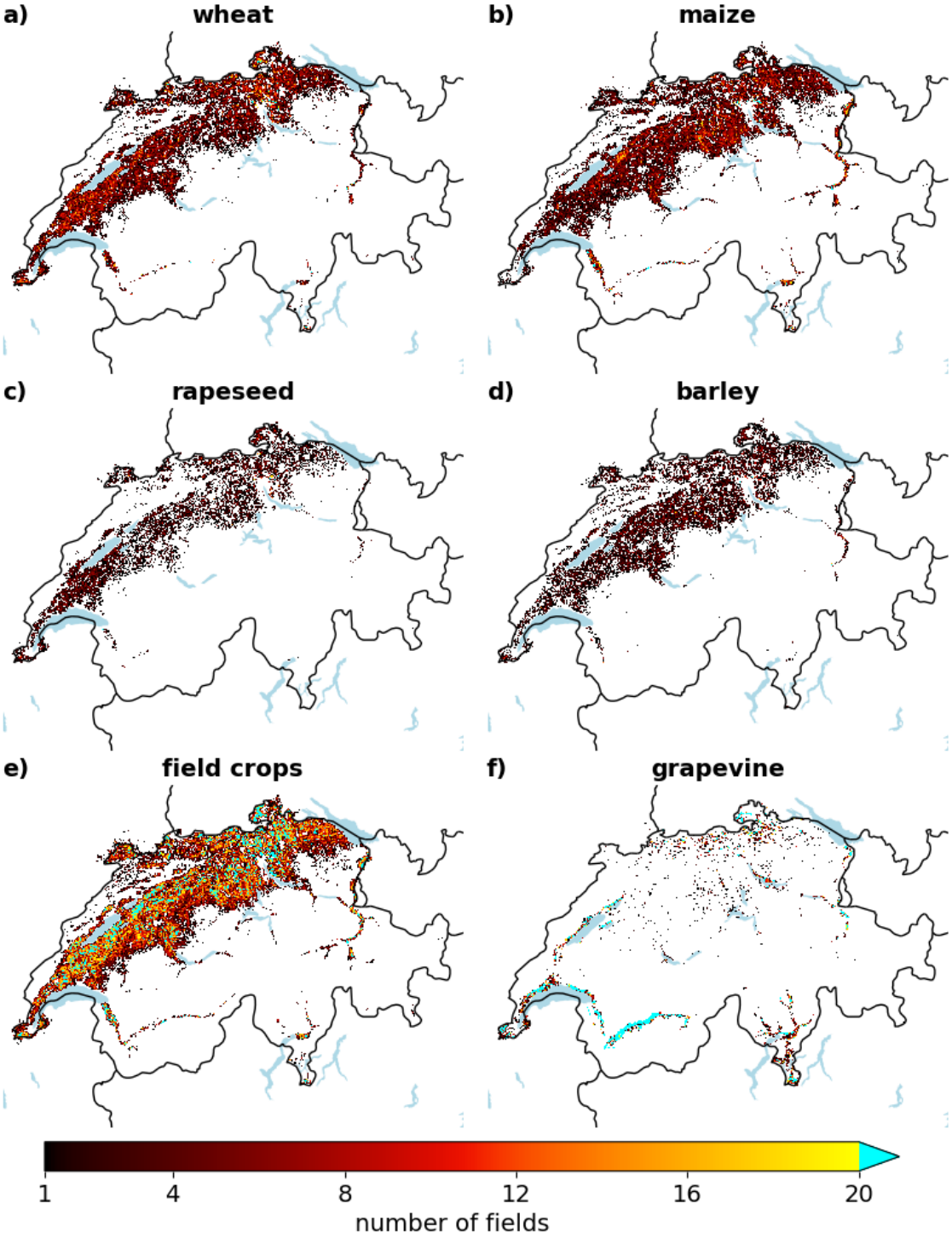

Figure A1. Cropland number density at 1km spatial resolution for (a) wheat, (b) maize, (c) rapeseed, (d) barley, (e) field crops, and (f) grapevine.

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

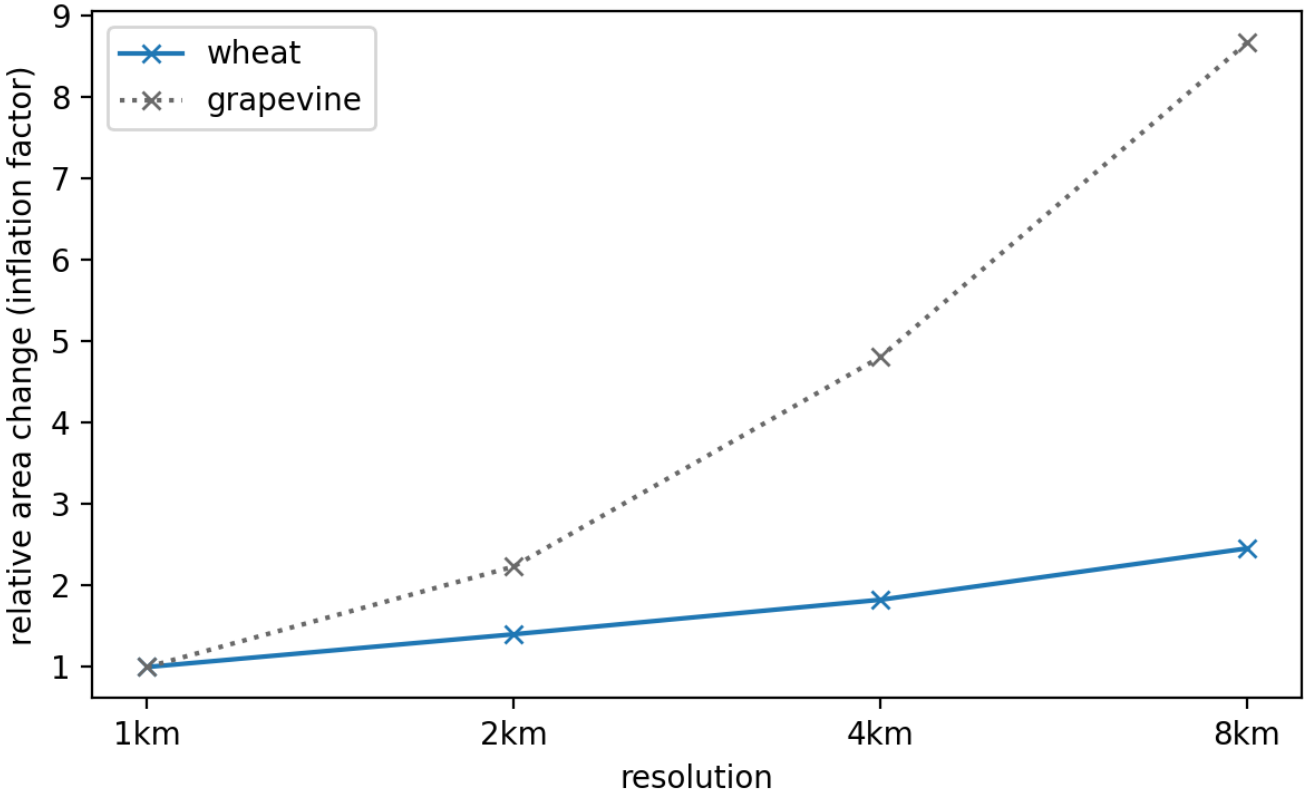

Figure A2. Relative change of the area covered by all exposure grid points (inflation factor) as a function of spatial resolution relative to the 1 km resolution for wheat (blue) and grapevine (grey, dashed).

https://doi.org/10.5194/egusphere-2023-2598
Preprint. Discussion started: 5 December 2023

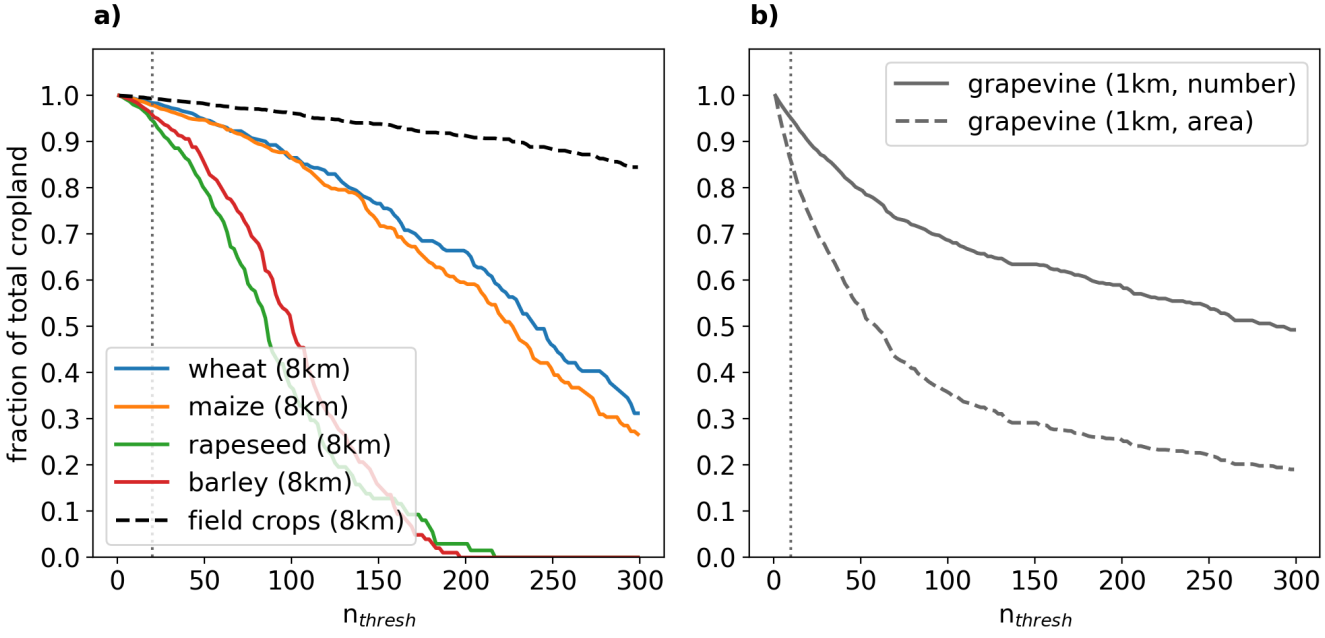

Figure A3. Change of the fraction of total number of fields included in the exposure as a function of n_{thresh} for (a) wheat (blue), maize (orange), rapeseed (green), barley (red), and field crops (black, dashed) at 8 km resolution and for (b) grapevine at 1 km resolution. In (b) the fraction of cropland area is also shown (dashed) because it deviates substantially from the fraction of fields for grapevine, but not crops shown in panel (a). The vertical bars denote the pragmatic choices of n_{thresh} (panel (a): 20, panel (b): 10) that avoid too high FAR but still includes a large fraction (>95%) of total exposed crop area / number of fields (see Fig. 6).