NHESS-2024-371

Reply document to referee comments 2 (RC2)

A modelled multi-decadal hailday time series for Switzerland

Lena Wilhelm, Cornelia Schwierz, Katharina Schröer, Mateusz Taszarek, and Olivia Martius

We thank the reviewer for their useful comments and the positive feedback. By addressing these comments, we can better clarify some crucial points and substantially improve the quality of the manuscript. The suggested changes are addressed in the following document. The comments of the reviewer are shown in black and our replies in blue. We number the specific comments for referencing purposes throughout the document (comment 1 = C1, etc.). Technical corrections are changed directly in the manuscript and are referenced with the starting line number, where removed parts are ~~crossed out~~ and new additions are in *italic*. All line numbers refer to the originally submitted manuscript.

Specific comments:

C1: line 28: You could add Manzato et al. (2022) here in addition to the Augenstein presentation. Their trend (Fig. 3) is not significant (but rather negative than positive) and at least it is published.

We thank the reviewer for this suggestion. The citation is added to the revised manuscript.

C2: 100: The intro is nicely structured and compact.

The authors are grateful for the positive feedback on the introduction.

C3: 91-94, 130-145, 162-169: I think there are some minor inconsistencies in your approach or how you explain it. You mention POH specifically includes small hail but then you use thresholds for POH and affected area which were tested for car insurance. To my knowledge, damage to structures and cars is dominated by severe hail (>2cm). Hence, to the reader it remains unclear what hail sizes the trend you find is representative of. It might be worth looking into the trend for more or less severe events (e.g., higher values of POH or larger areas?). For instance, it's good that you mention that the sensitivity to the area threshold was tested, but you could elaborate on this more and test other POH thresholds if feasible. I think such tests and discussions would improve the robustness of your results but I will leave it to you.

You are raising an interesting point. POH, per definition in the literature, captures hail of any size. However, preliminary comparisons with crowd-sourced hail size information do show some correlation between the POH value and hail stone size up to a POH approximately 80%. Further on, as you rightly point out, damages to cars occur typically only for hail stones that are larger than approximately 2cm. We include this information in the discussion.

Lines 89–94 are changed to:

This paper builds on the work of Madonna et al. (2018), however, here we increase the resolution of the analysis to daily, additionally include the South of Switzerland, and extend the time series back to 1959. As opposed to Battaglioli et al. (2023a) who used ESWD severe weather reports, ~~we use radar data to model daily hail occurrence of any size at the ground~~ *we use radar data as proxies to model hail day occurrence,* and we use an ensemble of two statistical models (logistic multiple regression and logistic generalized additive model (GAM)) leveraging the best-fitting predictors for each region individually. ~~We include smaller sizes as we work with the POH radar product, which is a proxy for hail of any size at the ground. Additionally, even small hail can be damaging to agricultural produce (Katz and Garcia, 1981).~~

Lines 126-145 are rewritten following a comment of reviewer 3:

In models classifying hail events, one typically selects the ERA-5 grid point that is temporally and spatially closest to the hail incident. However, for reconstructing past hail events, this is not possible because there is no information available on the hail event prior to the start of the observational period. Therefore, to model the occurrence of a hail day, we calculated ERA-5 profiles at grid points across the entire northern or southern domains at 12:00 UTC and then computed spatial averages of the values. The choice of using the value at 12:00 UTC, which exhibited the highest predictive skill, may be attributed to the fact that most storms in Switzerland occur in the afternoon. Thus, this value likely captures the atmospheric conditions conducive to storm formation.

Our definition of hail days focuses on days with more than just a single hail cell. The thresholds are set to capture events that in the past led to damages and affected somewhat larger areas

(Probability of Hail ≥ 80% over a minimum area of 580 km² for the northern domain and 499 km² for the southern domain, as detailed in Section 3.1).

Lines 162-169 are changed to:

To identify haildays in northern and southern Switzerland, this study uses daily POH data from 2002 to 2022 during the hail months of April to September. We here use the same domains and thresholds as Barras et al. (2021). The extent of the daily area of POH ≥ 80% is extracted separately for the domains north and south of the Alps (Fig. 1). *To qualify as a hailday, the daily maximum Probability of Hail (POH) must reach or exceed 80% over an area of at least 580 km² in the northern domain and 499 km² in the southern domain. Barras et al. (2021) determined that these thresholds correlate best with days when car damage was reported across Switzerland from 2002 to 2012.* ***This definition implies that hail large enough to cause damage to cars, approximately ~2cm in size, is considered.*** *The sensitivity of our model results to this threshold was tested by varying the area threshold, finding no significant impact on misses or false alarms, consistent with earlier studies indicating low sensitivity to area thresholds (Madonna et al., 2018).* ~~For a day to be categorized as a hailday, the daily maximum POH needed to equal or exceed 80% for at least 580 km2 in the northern domain and 499 km2 in the southern domain. Barras et al. (2021) found that these thresholds best correlate with days with car damage reported across Switzerland from 2002 to 2012. We tested the sensitivity of our model results to this threshold by varying the area threshold but did not find a significant impact on misses or false alarms, confirming earlier findings of low area threshold sensitivity (Madonna et al., 2018).~~ The findings regarding long-term trends, drivers of trends, and seasonality remain consistent regardless of the chosen POH area threshold, as anticipated, given our focus on haildays rather than specific hail sizes. Due to the length of the paper, we prefer not to delve further into the sensitivity analysis in the text.

C4: 235 and 299-307: I was wondering the whole time why no kinematic information is included. I would recommend adding a brief sentence here saying something like: „The lack of a kinematic predictor in the northern model will be discussed further later on."

We added a sentence to line 307:

[…] v_500 might indirectly represent wind shear. Wind shear, which is the change in wind speed and direction with height, is important for the organization and persistence of the storm system, specially for those capable of producing hail. A positive sign of v_500 indicates air moving towards the north at 500 hPa, which the model translates to higher probabilities for hail. This might also be related to the synoptic situation in the south of Switzerland, where moist, warm air is often advected from the Mediterranean (*Nisi et al. 2016,* Schemm et al., 2016). *The lack of a kinematic predictor in the northern model will be discussed further in section 6.1.* […]

C5: 290-296: Convective updrafts are not resolved in ERA5, so large OMEGA cannot be caused by this. In other words, I don't think this can be linked to the results of Lin and Kumjian (2022).

We thank the reviewer for commenting on this mistaken interpretation. We fully agree.

Lines 291-296 are changed to:

The vertically integrated vertical velocity (OMEGA_vint) denotes the vertical motion of air throughout the atmospheric column, primarily reflecting large-scale synoptic uplift or descent. The highest probabilities for hail occur when OMEGA_vint values are negative, signifying large-scale ascent. This atmospheric condition promotes the formation and maintenance of thunderstorms, thereby increasing *the likelihood of hail.* ~~The vertically integrated vertical velocity (OMEGA_vint) represents the vertical motion of air within the full column of the atmosphere. Negative values of OMEGA_vint indicate upward motion of air, which is crucial for the development of thunder- and hailstorms. Highly negative values of (OMEGA_vint) indicate very strong lifting, potentially with a very strong and narrow updraft. However, we do not see the highest probabilities for hail for those cases, but rather for median OMEGA_vint values (see 3). In the context of hail formation, this could mean that a less intense and wider updraft is more favourable than a very strong and narrow one, where hail embryos could be ejected prematurely, as already modelled by Lin and Kumjian (2022).~~

Lines 369-372 are changed to:

Similar to the vertically integrated vertical velocity OMEGA_vint, the vertical velocity at 500 hPa (w_500) is a measure for the vertical motion of air, here for the layer at 500 hPa. Negative values indicate *large-scale lifting* ~~and hence measure updraft strength.~~ The highest positive effect is achieved with the strongest negative vertical velocities.

C6: Also, while I agree that the role of synoptic lift is likely included in this predictor, the inclusion of omega is surprising because I would have expected orographic lift to dominate over Switzerland? Perhaps because the larger scale lift favors move widespread convection and hence more POH area it is still a useful predictor?

We agree that orographic lift is important for the initiation of deep moist convection in the Swiss Alps. While ERA5 data can capture some aspects of orographic lifting, its ability to accurately represent the small-scale mountain-valley circulations and convective initiation (with their connection to the diurnal cycle and differential heating) may be limited by its resolution. Since omega is the vertically integrated vertical velocity (from model level data), it combines information about the whole atmospheric column. Synoptic lift is included in this predictor, however, if extensive enough, orographic lift might also contribute to the integral. Schemm et al. (2016) also showed, that around 30-40% of Swiss hail storms form in a pre-frontal environment. However, in the northern logistic regression model the existence of fronts might already be captured by the Boyden Index (BI). The predictand OMEGA_vint had a high skill only in combination with moisture predictands, e.g. Q_vint, the vertically integrated specific humidity. So, the model might benefit from the combination of information on large-scale lift, fronts and moisture as well as instability.

C7: 362-364: Lin and Kumjian (2022) saw increasing hail potential until around 2500 J/kg. CAPE doesn't even reach such values in your Fig. 6 and the curve flattens at 500 J/kg already, so I don't think these results should be linked, at least not with further explanation.

Agreed, we removed "This relation has also been found by Lin and Kumjian (2022)" in line 363.

C8: 376-395 Some of your interpretations here were a bit confusing to me. I think this is a very active research topic but strong storm-relative winds have been shown to promote wider updrafts (Dennis and Kumjian 2017, Peters et al. 2020) and are hence important for hail. Also, weaker low-level winds have been suggested to be better, especially in north-south direction (Dennis and Kumjian 2017, Nixon et al. 2023). So your results or their interpretation are counter-intuitive, which should be clarified (you write that strong storm-realive winds are bad for hail but low-level shear good, at least that's how I understand your text, maybe you got them mixed up?). One explanation could be that the typical environments in Switzerland are different compared to these studies in the US. So the sensitivities to kinematic variables might be different. Which is worth to be discussed.

We thank the reviewer for this comment and will discuss this further before resubmission.

Lines 376-395 are adjusted:

Notably, the deep layer shear WS_06 exhibits a non-linear relationship to the response variable. For instance, WS_06 has its most negative effect at values around 0–10ms−1, transitioning to a positive effect from around 15ms−1. The curve’s slope declines noticeably in the presence of very high wind shear values, suggesting ~~again~~ that higher ~~bulk~~ shear does not always lead to a further increase in the probability of hail. Additionally, confidence intervals of the smoothing functions widen considerably towards the tails of each covariate distribution. In GAM‘s we are not limited by multicollinearity between model terms, which is why both WS_36 and WS_06 are in the northern model. The model seemed to prefer including both WS_36 and WS_06 over just one of them, as the individual predictor otherwise became non-significant and less important. Surprisingly, WS_36 has a negative linear relationship to hail. To gain a deeper understanding of how the model terms WS_36 and WS_06 interact, we further examined contour plots depicting conditional probabilities based on pairs of model predictors (not shown). The highest probabilities of hail are linked to high WS_06 but low WS_36 in the northern model. ~~Typically the difference in 3km to 6km is much smaller than the deep layer shear WS_06.~~ Hence, the model might indirectly learn, that the low-level winds (0-3km) are more important to hail growth (and storm dynamic, as the inflow dominates storm dynamic) than the 3-6km layer shear. *These findings are in contrast to those observed in modelling studies of hail in the US. High 3-6km wind shear influences the stretching of the hodograph, resulting in stronger storm relative winds. Strong storm-relative winds have been shown to promote wider updrafts and hail formation (Dennis and Kumjian 2017, Peters et al. 2020). Dennis and Kumjian (2017) and Nixon et al. (2023) found that weaker low-level winds are beneficial for hail formation, especially in north-south direction (Dennis and Kumjian 2017, Nixon et al. 2023). The southern logistic model, however, favored v_500 over any wind shear predictor. This discrepancy could stem from the unique environmental conditions in Switzerland compared to the idealized modelling studies conducted for individual hailstorms in the US. It's plausible that the sensitivities to kinematic variables differ between regions due to varying atmospheric dynamics and topographical features.*

~~With strong storm relative winds, the hail embryos might also be ejected out of the hail growth zone too quickly to experience significant growth (Dennis and Kumjian, 2017). This could explain why the model learns that high shear at this level leads to less probability of hail when differentiating between hail and no hail (and not hail size). We also need to mention, that we are just taking into account bulk shear which doesn´t include the curvature of the hodograph and the rotation of the storm.Dennis and Kumjian (2017) show that the biggest hail growth volume is achieved with a straight, long hodograph because of a strong and in the shear direction elongated updraft, rather than with a curved hodograph.~~

C9: 397: What do you mean by „circulate“? Most large hail seems to follow a single up-down trajectory while curving around the updraft (e.g., Kumjian and Lombardo (2017), Pounds et al (2023). No re-circulation with repeated ingestion into the updraft seems to happen. How this is in non-supercell storms is still unclear, but I don’t see a reason to assume differently.

Also, I’d suggest rephrasing to „deeper hail growth zone“ or „longer residence time in the hail growth zone“.

We thank the reviewer for their comment and agree with the misleading wording. We rewrite line 397 for clarity:

A lower freezing level suggests a greater potential for hail formation due to *a longer residence time of the embryo* in the hail growth zone. In this zone, hail embryos curve around the updraft, encountering areas of high supercooled liquid water, allowing them to accumulate and grow larger. ~~A lower freezing level may translate to a higher potential for hail formation because of the longer hail growth zone~~

~~where embryos are circulating through areas of high supercooled liquid water and are able to grow bigger.~~

C10: 398-403: Punge et al. (2023) also found that excluding freezing lvls<2400m helped reduce false hailstorm detections in higher elevation in South Africa. Might be relevant here.

We thank the reviewer for this citation. Lines 398-403 are changed to:

The highest probabilities of hail are achieved *for* ~~at~~ freezing levels between 2500 m.a.g.l. and 3500 m.a.g.l.. *Punge et al. (2023) also found that excluding freezing levels above 2400m.a.g.l. (and below 4845 m.a.g.l.) lead to a reduction of false hailstorm detections in higher elevation in South Africa when looking at characteristics of hail hazard based on satellite detection of convective storms.* A negative linear relationship is evident among freezing levels below 2500m.a.g.l., indicating that lower values of deg0l correspond to a reduction in hail probabilities and not an increase. This unambiguous relation has also been seen by Kunz (2007) before. This might suggest that our model does not learn about the melting or growing of the hail embryos from the freezing level*, but that it is associated with* ~~*connected to*~~ *the height of the cloud base and therefore the width of the updraft (Mulholland et al., 2021). The deg0l may also be a proxy for the surface temperature, as both are positively correlated (see Tables A1 and A2).*

C11: 470: I liked that you followed closely and compared your results to Raupach et al. (2023).

The authors appreciate this comment!

C12: Section 5.3: April and September don't have a good sample size and I don't see any increase for either month, but in the text you say "This leads to more events specifically at the beginning of the hail season (April-June)". Perhaps May-June would be more accurate?

Indeed, this sentence is misleading. We thank the author for his thorough review. Further examinations showed that the fraction of haildays occurring until a specific month did not show a stronger increase in April and September compared to May-June across different decades.

Lines 498-508 are changed to:

~~This section addresses the seasonal analysis of hail. Boxplots in Fig. 13 (a,b) show the modelled haildays per month for the whole time series period of 1959- 2022 for both domains. One can see a clear seasonal cycle of hailday occurrence. There is strong year-to-year variability in the monthly number of haildays (see e.g. circles for month July). In Fig. 13 (c,d) the mean number of haildays per month is plotted for each decade in differently colored curves. Decade 1960´s includes the years 1960-1969. The years 1959 and 2020-2022 are excluded, as we wanted to have the same number of years in each decade. In the last two decades (blue and purple curve), there is a strong increase in haildays. This leads to more events specifically at the beginning of the hail season (April-June) and to a shift of the peak of the convective season towards earlier months for both regions. These results are also evident when looking at the seasonal cycle by week or day of the year. Due to the large variability, the differences in the monthly curves are not significant and no systematic shift can be seen (see also cdf plots in Fig. A3). However, as our analysis was limited to the months of April to September, we cannot make assumptions regarding potential changes in hail events in the months preceding or following the modelled period.~~

This section addresses the seasonal analysis of hail occurrence. Boxplots in Fig. 13 (a,b) illustrate the distribution of modelled hail days per month over the entire time series from 1959 to 2022 for both *domains, revealing a pronounced seasonal cycle in hail day occurrence. The monthly number of haildays exhibits considerable year-to-year variability, as evidenced by the spread of the circles, particularly notable in July. Fig. 13 (c,d) further portrays the mean number of hail days per month across different decades, with the exclusion of years 1959 and 2020-2022 to ensure consistency in the number of years per decade.* ***Notably, the last two decades (blue and purple curves) exhibit a marked increase in hail days, which is strongest in May and June. While the monthly curves display considerable variability, their difference is not significant and no systematic shift is evident, as illustrated by cumulative distribution function (cdf) plots in Fig. A3.*** *However, as our analysis is confined to the months of April to September, we refrain from making assumptions regarding potential changes in hail events preceding or following the modelled period.*

C13: 525-549: I agree with these explanations, but are your hail day thresholds appropriate then, since they are trained with vehicle damages and hence larger hail ?

We thank the reviewer for their comment. Following comment C3, we will remove our assertion about differentiating between any hail and large hail. However, the low feature importance and low significance of shear in nearly all tested models shows that in our study area, high shear might not be as important for hailstorm formation.

Lines 532-546 are rewritten:

This leads us to our second point, which is that *high* wind shear *might be less important for hailstorm formation in complex terrain.* ~~more important for distinguishing between small and large hail (< 2cm) than for the prediction of hail occurrence vs. no hail.~~ Several studies (Brooks et al., 2003; Kaltenböck et al., 2009; Púˇcik et al., 2015; Taszarek et al., 2020a) have highlighted the importance of wind shear in the formation of hail, as it influences the lifetime and structure of a storm. While large shear values are required to form supercells, which are more likely to produce hail, hail also develops in lower-shear environments (Blair et al., 2021; Kumjian and Lombardo, 2020). ~~Additionally, many of those studies only focussed on large hail (> 2cm or > 5cm).~~ ~~In fact, Markowski and Richardson (2010) and Dennis and Kumjian (2017) showed that wind shear primarily drives storm type and hail diameter.~~ Houze et al. (1993) found that a significant fraction of Swiss hailstorms are not supercells but ordinary and intermediate-type storms. Punge and Kunz (2016) revealed that in complex terrain, high wind shear might not be necessary for hailstorm formation. Hail events in low-shear environments can be explained by a proximity of mountain ranges where environmental wind shear is enhanced by an interaction of the wind field with orography, which is often the case for the Alps (Kunz et al., 2018). Lastly, with such complex terrain, shear might be driven by local conditions (e.g. Alpine pumping) that are too small to be resolved by ERA-5´s resolution. ~~ERA-5 could show weak shear, but due to orographical enhancement, storms actually benefit from larger shear values.~~ This is something that can be better resolved with high-resolution convection permitting models (with approx. 1x1km spatial resolution).

C14: 600: Raupach et al. had to cover a much larger region and different climate zones. Perhaps it is fair to mention that the skill could be linked to such differences?

We appreciate the importance of recognizing such distinctions and aim to maintain humility in our interpretation of the models. Lines 599-602 are changed to:

Our models outperformed those mentioned in Raupach et al. (2023a) due to lower FAR and higher HSS values. For the different studies, HSS ranges from 0.1 to 0.4 compared to our models' HSS of 0.73 (north) and 0.35 (south). FAR ranges from 0.57 to 0.8 compared to 0.23 (north) and 0.35 (south) for our models. *However, it is important to note that Raupach et al. (2023a) conducted their study over a much larger geographic area encompassing diverse climate zones. Other studies have demonstrated* *comparable performance to ours, such as Battaglioli et al. (2023a) using ESWD hail reports and ERA-5 data, López et al. (2007) utilizing radar and radiosonde data and Gascón et al. (2015) using severe storms reports and WRF vertical profiles.* ~~*However, there are also studies showing comparable performance skills, like Battaglioli et al. (2023a) using ESWD hail reports and ERA-5 data, López et al. (2007) using radar and radiosonde data and Gascón et al. (2015) using severe storms reports and WRF vertical profiles.*~~

C15: 647 and abstract: I don't see a longer peak in Fig. 13. (see also comment on September above)

We agree. Lines 646-648 are changed to:

Furthermore, we can see an increase in haildays in the last two decades, which is strongest in May and June. ~~and a slightly earlier and longer peak of the hail season compared to earlier decades. Still, there is~~ *However, there is no clear shift of the seasonal cycle towards an earlier start and earlier end, and differences in monthly distributions across decades are not significant.*

Line 13-15 are changed to:

In the last two decades, we can see an increase in haildays, *which is strongest in May and June*. ~~at the beginning of the hail season and an earlier and longer peak, however,~~ However, there is no systematic shift in the seasonal cycle *across decades.*

C16: Overall, I liked your thorough discussion and conclusions.

We are grateful for the reviewer's positive feedback.

C17: Acknowledgements: You mention that parameters from thundeR were tested. Maybe I missed this in the text, but some discussion of it might be insightful, no?

We conducted tests on thundeR parameters for two reasons:

1.) Some of our parameters were derived from ERA-5 data at a lower resolution (0.5x0.5°) rather than the highest available resolution (0.25x0.25°). We aimed to determine whether this resolution difference influenced model performance and whether calculating profiles based on e.g. the pixel with the highest CAPE in space and time would enhance performance compared to our method of spatially averaging profiles. However, neither scenario resulted in improved performance.

2.) We wanted to investigate if incorporating additional convective parameters, initially not included in the parameter selection (e.g., MUCAPE above -10°C), would significantly enhance model performance. However, our findings did not support this. It's worth noting, however, that we did not dedicate the same level of effort to train and refine these models as we did for our original selections. It's plausible that with extensive further work, models with similar or better performance could be identified. We believe that ongoing research and the availability of more and higher resolution data in the future could further improve such models.

Due to the manuscript length and potential reader confusion, we have opted not to include this discussion.

C18: Also, since M. Taszarek is in the author list, I'm not sure if it's necessary to acknowledge his contribution. Your choice though.

Thank you for bringing this to our attention.

Technical Corrections:

L 27: I'm not sure if this is the standard formatting in egusphere but with some references you use brackets around the years (when the reference starts wiht something line „e.g.") and in some not. Also, it is common to put a comma after „e.g." I think.

Thank you for this comment. All references and the e.g. are checked again in the revised manuscript.

L 43: Just a minor thing, but I think „ERA5" is more commonly used? Adjusted in all sections.

L 48: I suggest removing „a" before „sufficient" Adjusted.

L 68: I suggest writing either "both thermodynamic" in the brackets or "...can be grouped into thermodynamics (instability and moisture) and kinematic conditions." to make clearer that „thermodynamic" refers to both instability and moisture

Thank you for this suggestion, we changed line 68 to:

The variables can be grouped into the categories of instability, moisture (*both* thermodynamic) and kinematic conditions.

L 195: Suggest removing extra brackets in the exponent. Adjusted in all sections.

Fig. 3: Units are missing here in some of the following figures. Will be added in the revised manuscript.

L 455: Add a space before „respectively". Adjusted.

L 564: „choose" Adjusted.

The references to Augenstein et al. (2023) and Mohr et al. (2015 a,b) need some fixing.

Manzato, A., S. Serafin, M. M. Miglietta, D. Kirshbaum, and W. Schulz, 2022: A Pan-Alpine Climatology of Lightning and Convective Initiation. *Mon. Wea. Rev.*, **150**, 2213–2230, https://doi.org/10.1175/MWR-D-21-0149.1.

Nixon, C. J., J. T. Allen, and M. Taszarek, 2023: Hodographs and Skew Ts of Hail-Producing Storms. *Wea. Forecasting*, **38**, 2217–2236, https://doi.org/10.1175/WAF-D-23-0031.1.

Peters, J. M., C. J. Nowotarski, J. P. Mulholland, and R. L. Thompson, 2020: The Influences of Effective Inflow Layer Streamwise Vorticity and Storm-Relative Flow on Supercell Updraft Properties. *J. Atmos. Sci.*, **77**, 3033–3057, https://doi.org/10.1175/JAS-D-19-0355.1.

Punge, H. J., Bedka, K. M., Kunz, M., Bang, S. D., and Itterly, K. F.: Characteristics of hail hazard in South Africa based on satellite detection of convective storms, Nat. Hazards Earth Syst. Sci., 23, 1549–1576, https://doi.org/10.5194/nhess-23-1549-2023, 2023.

We thank the reviewer for his thorough proofreading. All references are fixed.