NHESS-2024-371

Reply document to referee comments 1 (RC1)

A modelled multi-decadal hailday time series for Switzerland

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

The authors thank the reviewer for taking the time to review this manuscript so thoroughly. The constructive feedback and useful comments showed us where we need to clarify points. The suggested changes substantially improve the manuscript, and we addressed all comments in the following document. The comments of the reviewer are shown in black and our replies in blue. We number major comments for referencing purposes throughout the document (comment 1 = C1, etc.). Other comments are addressed by making changes in the manuscript directly. Removed parts are ~~crossed-out~~ and new additions are in *italic*. All line numbers refer to the originally submitted manuscript.

Major comments:

C1: Some of the indices used in this manuscript are quite outdated and have not been used in state-of-the-art meteorological research since the 1970s (e.g., LI, TT, BI, K index). [This list is just a partial list of the ones that are outdated.] These indices are combinations of other variables that are likely in the analysis anyway. (If not, they could easily be replaced with dewpoint at 850 hPa or air temperature at 700 hPa.) This manuscript cannot claim to be using modern methods with such outdated meteorological quantities. I recommend that the authors remove such outdated indices and perform the analysis again. Doing so will improve your paper and give you credibility. This is a potentially state-of-the-art paper, but other meteorologists who read this manuscript will struggle to accept it with such outdated metrics in it.

Doswell and Schultz (2006) have more on the history and inappropriateness of these indices.

Doswell, C. A. III, and D. M. Schultz, 2006: On the use of indices and parameters in forecasting severe storms. Electronic J. Severe Storms Meteor., 1(3), 1–22, https://ejssm.com/ojs/index.php/site/article/view/4.

Thank you for your feedback. We acknowledge your concern regarding the use of certain convective indices, and we appreciate the opportunity to address this issue.

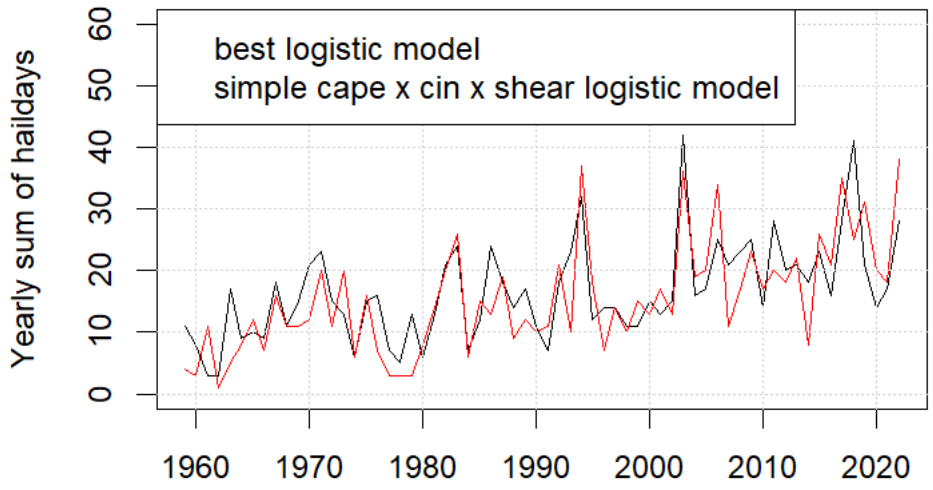
Our study aimed to be comprehensive by including various parameters, both traditional and modern, to statistically model hail occurrence in a region with complex terrain. We also included widely accepted parameters like CAPE, shear, and the individual ingredients of the composite indices (refer to Table A2). Additionally, we tested 132 convective parameters from the thundeR package. The model selected the variables independently based on minimizing a loss function. The models presented in our manuscript yielded the best performance.

We agree with Doswell and Schultz (2006) arguing that all diagnostic variables are subject to errors, some of them more volatile than others. We need to stress, however, that our goal was not to forecast individual hail events perfectly, but rather to achieve the best possible reconstruction of hail days over a broad study area with complex topography. We think that the high volatility of “older” parameters is counterbalanced by the combination of multiple variables in the same model, adding to a more complete picture of the atmospheric profile.

While we agree that not all convective parameters in our selection are prevalent in recent literature, we respectfully disagree with the assertion of their outdatedness. We have included citations from the last two decades that utilize these parameters (see Section 4.1, lines 271, 277, 282, and Section 6.2, lines 586-592). We also transparently discussed the limitations of “older” convective parameters in Section 6.1. Furthermore, recent presentations at the 4th European Hail Workshop (March 5-7, 2024) demonstrated that parameters like TT are still in use and effective in hail modeling studies, as evidenced by e.g. Boris Blanc's and Luis Ackermann´s work.

It is crucial to distinguish between the predictive skill of a single variable in a univariate model and the role of an individual variable in a multivariate model. In our models the combination of TT and SLI had great predictive skill, however, if their skill was tested by themselves, they indeed performed worse than considering e.g. just CAPE. Even with such “simple” statistical models, comprehending the entirety of what patterns and relationships between parameters the model learns is not fully transparent, especially when composite parameters are in play. In this regard we again agree with Doswell and Schultz (2006).

Lastly, we need to mention that the results concerning trends, drivers of trends, and seasonality are robust to the selection of the model. The following plot shows a comparison of the yearly number of haildays for the study area north of the Alps between our best logistic model (see Sect. 4.1, line 235) with a "simpler" model using only MUCAPE, CIN and wind shear from 0-6km. The trends in yearly haildays are very similar. MannKendall trend tests show that the tau values are 0.36 and 0.38 comparing the best to the "simple" model, indicating only marginal difference in the strength of the long-term trend.



Additionaly, we compared the yearly number of hail days in southern Switzerland with those in northern Italy as modelled by Battaglioli et al. (2023) (overlapping study area). Despite different models and parameter selection, there was a strong overlap in findings, reinforcing the reliability of our results.

We appreciate your feedback and believe that our comprehensive approach and rigorous validation steps maintain the credibility and relevance of our study. We remain open to incorporating further suggestions and refining our analysis to enhance the robustness and acceptance of our findings.

C2: The authors conflate the various indices, not properly understanding their original intent. For example, at lines 48–49, why "hailstorm"? These ingredients apply to any organized deep moist convective storm (Doswell et al. 1996). The ingredients for hail to be produced from organized deep convective storms are in addition to these. Please clarify. See also line 54.

Doswell, C. A., III, H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients based methodology. Wea. Forecasting, 11, 560– 581.

The authors thank the reviewer for this comment. Indeed, we should differentiate the ingredients necessary for the formation of deep moist convection (DMC) from those needed for hailstone formation.

Lines 48-54 are changed to:
~~A combination of multiple mesoscale variables is required to estimate the hail potential of the atmosphere.~~ *The development of deep moist convection requires an unstable atmosphere, sufficient low-level moisture, vertical wind shear and an initiation mechanism (e.g. Doswell et al. 1996, Johns and*

Doswell (1992). For hailstones to form in a storm, in addition three main ingredients are needed: an embryo particle, typically graupel or frozen drops, which serves as the nucleus; an abundance of supercooled liquid water collected through riming and accretion; and sufficient time within the storm's updraft for the hailstone to grow, determined by its trajectory and the updraft's characteristics (Allen et al 2019, Kumjian et al. (2021), Kumjian and Lombardo (2020)).
Several studies have shown that both mesoscale and synoptic flow-conditions, including the presence of fronts, and their interaction with complex terrain are significant factors influencing the development of hail storms ~~driving hail formation~~ in Europe (e.g. Brooks et al. (2003, 2007); Cacciamani et al. (1995); Costa et al. (2001); Giaiotti et al. (2003); Kunz and Puskeiler (2010); Piasecki et al. (2023), Schemm et al. (2016)). Regional characteristics, such as terrain barriers, local wind systems, or warm water surfaces influence the relative importance of the four components necessary for hailstorm development ~~formation~~, which is why this study looks at the region north and south of the Alps separately.

Furthermore, many of the indices used are not relevant to hail production in storms. Why include them if not relevant?

In our models, we don't specifically differentiate between deep moist convection and hail; rather, the models learn about both simultaneously. Using low-resolution ERA5 data, indices specific to hail production in storms may not exist, as they would require detailed information on hail embryos. Instead, we focus on variables capturing atmospheric potential for deep moist convection and supplemented with factors essential for large hailstone formation, such as vertical shear and storm organization or information on the freezing level. The model finds a relation to hail day occurrence by using the multi-variate proxy.
We anticipate that ongoing research and access to higher-resolution data in the future could enhance these models further.

C3: Line 290–291: OMEGA_vint is the vertical velocity on the scale of the ERA-5 (e.g., 25-km grid spacing). So, it cannot represent the vertical motion on the scale of the convective storms (Doswell and Bosart 2001). And, it is certainly not representing the ejection of hail embryos, as is implied in the manuscript.

Doswell, C.A., Bosart, L.F. (2001). Extratropical Synoptic-Scale Processes and Severe Convection. In: Doswell, C.A. (eds) Severe Convective Storms. Meteorological Monographs. American Meteorological Society, Boston, MA. https://doi.org/10.1007/978-1-935704-06-5_2

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).~~

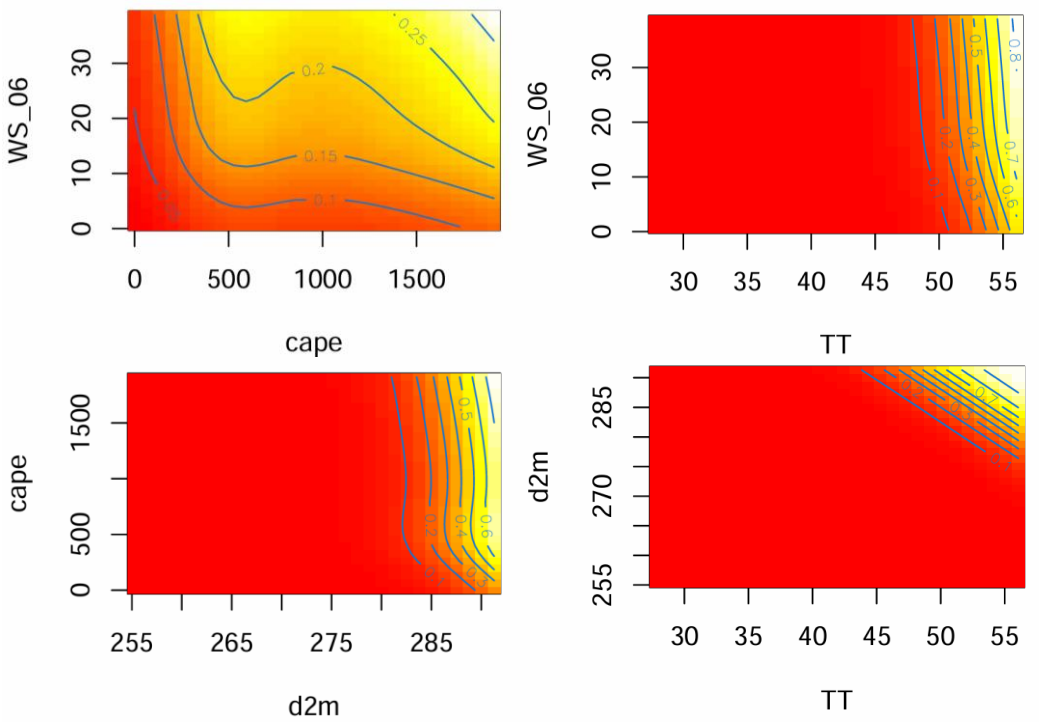
Lines 369-372 are changed to:

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 an upward motion ~~and hence measure updraft strength.~~ The highest positive effect is achieved with the strongest negative vertical velocities.

C4: I have concerns about how the authors are using the various variables. I don't get a sense that there is an attempt to understand the physical justification for them or even appropriateness of them, as the previous comments have indicated. Instead, the paper is excessively couched in terms of statistics. It appears to be one of the reasons that these outdated convective indices are used in this paper.
As an example of this, I have the following comment regarding text at lines 562–564. If I understand the authors' argument here correctly, they are happy using outdated indices and not understanding why they work, as long as they get a time series they can work with. Do I understand that correctly?

We thank the reviewer for their constructive feedback. We appreciate the opportunity to address the concerns regarding our use of various variables and the perceived emphasis on statistical methods over physical justification.

Regarding the specific text at lines 562-564, we did not intend to imply that we are content with using indices without understanding their mechanisms. Rather, we aimed to highlight that our data-driven approach allowed the models to select the best variables for our specific goals and location, even if those variables included some less common indices. This approach does not negate the need for physical justification; instead, it complements it by leveraging statistical strength to handle data limitations, which is often overlooked. While we tried to understand each predictor's role (see Fg. 3-6 and Sections 4.1, 4.2, 6.1), there are inherent limitations to what we can infer about the model's learning process. As an example, we are providing additional analyses performed to explore the interactions between model parameters in one model. The heatmaps below show how hailday probabilities vary with specific combinations of predictor values in the southern logistic GAM model. Since we cannot grasp more than two to three-dimensional interactions, we cannot understand what the model learns from the combination of 5 (composite) predictors fully. In this regard we acknowledge that our interpretations are just theories.



We changed lines 562-564 for clarity:

We do not claim that the combination of LI and TT are better than e.g. CAPE and shear in *forecasting individual hail cells or* differentiating between no hail, hail and large hail. Rather we want to make clear, that our ~~synergy~~ specific combinations of around five variables in the ~~5D~~ models worked best for *the reconstruction of haildays in the Swiss study areas using the POH radar proxy and low-resolution ERA-5 data. While our data-driven approach identified some less common indices, the statistical models leverage these indices effectively within the constraints of our data, and this statistical approach complements our physical understanding.* ~~our exact goal and location, namely, reconstructing Swiss haildays in the last 70 years with the POH radar proxy and low-resolution ERA-5 data. We assume that the models learn how to best deal with the limitations of the data sets and chooses the best variables for our specific questions in a data-driven approach. We did not expect to learn about new processes from the models, but the models seem to agree with previous knowledge. We chose ERA-5, as this is the best available product for multidecadal analyses.~~ That said, one should not transfer our models to the future or to other regions around the globe. For forecasting applications the COSMO/ICON 1x1km analyses are much better suited *than ERA5.*

C5: There appears to be inconsistency in the messaging about how to interpret the trends. For example, line 618: If “it is not feasible to directly extrapolate our modelled trends into reality”, then of what value is this study?
How is this statement consistent with your stated goal of building “a new multidecadal daily hail time series” (line 635)?
Or with, “With this time series we wanted to analyse long-term trends and changes in frequency, seasonality and the variability of model-derived Swiss hailstorms in the past decades.” (Lines 640–642)?
Or with, “The final ensemble model reproduces the interannual variability and seasonality of the hail proxies well. The reconstructed time series shows a strong significant positive trend in the number of yearly haildays in both regions from 1959-2022. The trend is also significant and positive when looking at the period of 1979-2022.” (Lines 643–645)?
How does the reader interpret the different messages from these different sentences?

Thank you for pointing out the inconsistencies in our messaging about interpreting the trends. We understand the importance of clear communication and will make necessary adjustments.

Our intention was to highlight the inherent uncertainties in using convective parameters from ERA5 data instead of direct observations when analyzing hail trends. Specifically, the variations in trends observed in similar modelling studies presented at the recent European Hail Workshop underscore the need for caution and further investigation before drawing definitive conclusions.

To limit reader confusion, we will remove or adjust the following statements:

Two sentences in Line 617-619 are removed:
~~Hence, it is not feasible to directly extrapolate our modelled trends into reality. To achieve this, a more extensive period of observational data would be required.~~

Lines 640-642 are changed:
~~With this time series we wanted to~~ This time series was used to analyse long-term trends and changes in frequency, seasonality and the variability of model-derived Swiss hailstorms in the past decades.

Lines 650-653 are adjusted:
The main purpose of this study is to offer an alternative framework to study intra-annual variability, trends, and changes in the seasonality ~~of Swiss hail occurrence~~ in the past without long-term direct hail observations. ~~Our goal was not to predict every hail event in Switzerland per grid cell, but to reconstruct strong hailyears in the past.~~

C6: The paper is in severe need of proofreading, as indicated by the large number of minor comments, grammatical errors, and inconsistencies noted throughout the manuscript. A note to the authors: Many of these issues could have been fixed with more care spent proofreading the manuscript, which I encourage the authors to do for this and subsequent submissions to journals. Peer reviewers should not have to identify these large numbers of issues to improve submissions. Authors should take their own responsibility for the quality and professionalism of their submissions.

We apologize for the numerous errors and inconsistencies in the manuscript and appreciate the reviewer's effort in identifying them. Despite multiple rounds of proofreading, it is clear that we missed several issues due to the manuscript's length. We will ensure thorough proofreading before resubmitting.

Other comments:
We thank the reviewer for their comments. The authors will adjust all the mentioned technical corrections in the revised manuscript before resubmitting.

1. Title: The authors may want to be more specific about what kind of "model" is used and what "multidecadal" means (i.e., what years were used). We will add the study period in brackets "(1959-2022)" to the title to provide a clear temporal context. The authors believe adding more details about the model to the title might make it unnecessarily lengthy and complex. Instead, we have ensured that the nature of the model is clearly described in the abstract, allowing readers to quickly understand the type of model used upon reviewing the paper.

2. Throughout the manuscript: Hyphenate "hail-day" when modifying another word (e.g., "hail-day time series", but "the number of hail days"). Adjusted throughout the manuscript.

3. Line 7 and throughout: Use an en dash to separate two years, not a hyphen and spaces. See also lines 78, 149, 199, and 200, for example. Fix throughout. (Note that you do it correctly at line 79.) Adjusted throughout the manuscript.

4. Line 9: GAM is not defined. Definition is added.

5. Lines 13 and 15: Delete "we can see" and reword the sentence. Same with "we can now study". Lines 13-15 have been changed to:
In the last two decades, there has been a considerable increase in haildays, which is strongest in May and June. There is no systematic shift in the seasonal cycle across decades. This time series allows us to study the local and remote drivers of the interannual variability and seasonality of Swiss hail occurrence.

6. Line 14: "however" cannot be used as a conjunction. See also line 87.
https://www.iup.edu/writingcenter/writing-resources/grammar/common-problems-with-however therefore-and-similar-words.html
Unfortunately, the website link is no longer available, but we acknowledge that "however" cannot be used as a conjunction and will change any misuse in the manuscript. See previous comment for changes.

7. Line 21: "one of the most complicated meteorological phenomena": I'm not sure that this is appropriate. Please reword. Line 21f changed to: *Addressing hail hazards is challenging, as hail is associated with severe thunderstorms, a complex meteorological phenomenon involving intricate physical processes. Thunderstorms feature a small spatial scale, vigorous development, and interactions ranging from synoptic to microphysical spatial dimensions.*

8. Lines 31–32: This sentence is a tautology: "With hail hotspots in Switzerland, the Alps are affect by hailstorms in Switzerland." Lines 31-32 changed to: *The pre-Alpine regions north and south of the Alps are regularly affected by severe hailstorms (Fluck et al. 2021, Nisi et al. 2016). In Switzerland hail frequency hotspots are located in the Jura, the Entlebuch, and in Ticino, southern Switzerland (Nisi et al., 2016).*

9. Line 47: Why "mesoscale"? We wrote mesoscale because ERA5 cannot capture the microscale processes. We agree that this sentence is misleading and removed it in line 47 following comment C2.

10. Line 51: I am uncomfortable with the word "driving" here. Synoptic and mesoscale conditions often set the stage for convective storms that may or may not produce hail. So, "associated" would be a more neutral and accurate word to use in this sentence. We agree. Line 51 is changed to: *Several studies have shown that both mesoscale and synoptic flow-conditions, including the presence of fronts, and their interaction with complex terrain are often associated with hail formation in Europe (e.g. Brooks et al. (2003, 2007); Cacciamani et al. (1995); Costa et al. (2001); Giaiotti et al. (2003); Kunz and Puskeiler (2010); Piasecki et al. (2023)).*

11. Line 55: Why is "(or north)" in parentheses? Either delete the parentheses or delete the parenthetical. See also lines 61 and 249. Fix elsewhere in the manuscript.
We removed the parenthesis around "(north)" in lines 55, around "(hail producing)" in line 61 and removed "~~(highly)~~" in line 249. We will check throughout the manuscript.

12. Line 57: Air masses are transported, not advected. Quantities such as temperature can be advected. Write precisely. We thank the reviewer for this suggestion and want to write as clearly as possible. Line 57 "~~advected~~" is changed to "transported".
13. Lines 68–71: Except for CAPE, these indices are outdated. We refer to comment C1.

14. Lines 78–79. This sentence is poorly written and unclear. Delete the comma after "found". Lines 78-79 are rewritten: They found ~~that the potential for hail events will increase in the future (2021–2050) compared to the past (1971–2000), but only statistically significant in the northwest and south of Germany.~~ *, that in the future (2021–2050), the potential for hail events is projected to increase compared to the past (1971–2000), with statistically significant rises observed exclusively in the northwest and south of Germany.*

15. Line 85: "Poisson" should be capitalized. Adjusted.

16. Line 106: Why is EchoTop capitalized? The "Z" in dBZ should be capitalized and italicized, as per convention. Changed throughout the manuscript.

17. Lines 108, 122, and 123: Change "resolution" to "grid spacing" because these two terms are not equivalent. Adjusted.

18. Line 114: I don't understand the phrase "area up to 140 km". Units are not consistent. Why up? Reword to be more clear. We agree and rewrote lines 114f: We ~~here~~ use thoroughly quality-checked and reprocessed POH data ~~of~~ *from* the recently published Swiss hail climatology (Trefalt et al., 2023; Schröer et al., 2023) and ~~consider the area up to 140km around the five radar stations (see Fig. 1) to minimize planar artifacts and ground clutter~~ *consider areas within a 140 km radius around the five radar stations (Fig. 1). This limitation helps minimize planar artifacts and ground clutter.*

19. Lines 114, 125, 179, and 202: Delete "see". It is unnecessary as the citation to the figure number is sufficient. Fix throughout. Adjusted throughout.

20. Line 136 and throughout: Are colons used in UTC time? Generally not that I've seen. Delete the colons throughout. Adjusted throughout.

21. Line 150: I don't know what "sereal" means. If you mean "serial", I still don't know what that means in the context of this sentence. Please reword. The authors meant to write "several". We will remove the word, as it is unnecessary.

22. Line 154: Delete "It is important to note that", which is unnecessary. The sentence is stronger without this phrase. Deleted.

23. Line 161: There should be a brief introductory paragraph after section 3, but before section 3.1, describing what will be discussed in section 3. Give the reader some context for what follows. The same thing is true for section 4. Agreed, we will add short introductory paragraphs to the revised manuscript.

24. Line 173: I would argue that "mesoscale" should be replaced with "microscale". Agreed, mesoscale is changed to microscale.

25. Line 176: Is that 27.0 hail days? It should be consistent with the number of decimal places in the other number. Adjusted.

26. Line 182: No need for "as shown in". Just put the citation in parentheses at the end of the sentence. Adjusted.

27. Figure 1: There are stray characters "1e6" in the lower-right corner and top left. I don't understand why it is there. Make its purpose more clear or delete. The stray characters "1e6" in the lower-right corner and top left are there to indicate multiplication by 1,000,000. This notation is used

to avoid adding six zeros to each coordinate number. Switzerland's official coordinates are based on the LV95 reference frame, with the origin point located in Bern. The coordinates of this point are E = 2,600,000 m and N = 1,200,000 m. We use the "1e6" notation for simplicity.

28. Line 216: Extra space between "hail days" and the period. Delete. Deleted.

29. Line 245: Avoid sentences that start with "this <verb>". The reader often does not know what "this" refers to. Please rewrite to be more specific. Line 245 was changed: This *difference in performance metrics* suggests that the northern model can overall better distinguish between hail and no haildays and that it misses fewer haildays than the southern model.

30. Line 248: Italicize p, as well as other variables throughout the manuscript. Adjusted throughout the manuscript.

31. Line 249: What is "insignificant". It is unclear. Avoid misplaced modifying phrases.
https://opentextbc.ca/advancedenglish/chapter/misplaced-and-dangling-modifiers/
We agree and thank the reviewer for reminding us of the correct use of modifiers. We rewrite line 249: *The model´s performance decreased marginally when removing the factor month.*

32. Line 261: Similar problem with "this indicates".
Line 261 changed to: This *overlap in the curves* indicates that all predictors are well fitted and do not need further modification

33. Line 294: What is "3"? We changed "(see 3)" to "(Fig. 3)".

34. Lines 294, 316: What does "this" refer to? We appreciate the reviewer's attention to detail in identifying errors in the manuscript.
We rewrote lines 291-296 following comment C3.
Lines 316-317 are rewritten for clarity: ~~When applying the combination of variables of the model north on the region south and vice versa, the coefficients changed and the predictive skill declined. This again justifies the use of an individual combination of predictors for each region, rather than applying one model for the whole area of Switzerland.~~ *When the variable combinations from the northern model were applied to the southern region and vice versa, the coefficients changed, leading to a decline in predictive skill. This difference in coefficients and predictive skill reinforces the necessity of using unique sets of predictors for each region instead of a singular model across all of Switzerland.*

35. Line 296: Delete "already". Deleted.

36. Line 297: Avoid one-sentence paragraphs.
https://warwick.ac.uk/fac/soc/al/globalpad
rip/openhouse/academicenglishskills/writing/paragraphing/
Lines 297-298 are rewritten: ~~The vertically integrated specific humdidity (Q_vint) reflects the total amount of water vapor available for hailstorms to develop and hence, the larger Q_vint, the higher the probability of a hailstorm.~~ The vertically integrated specific humidity (Q_vint) quantifies the total amount of water vapor available *in the atmospheric column, serving as an indicator of the moisture available for hailstorm development. Consequently, a higher Q_vint translates to an increased likelihood of hail occurrences.*

37. Line 311: Why is "recursive" capitalized? Adjusted.

38. Line 311: "e.g." is Latin for "for example". As such, there is no need for "etc.", too. "etc" is removed.

39. Line 313: Delete the two "see"s. Deleted.

40. Lines 314, 330, and 359: "However" is used as a conjunction. Revise.
Lines 308-317 were rewritten to enhance readability:
All the mentioned connections are part of a complex interplay of atmospheric conditions that contribute to hailstorm development. This is why we examine a combination of various parameters, ~~including those related to instability, moisture, and wind patterns,~~ to assess the likelihood of hailstorms in a synergy in our models. *When the variable combinations from the northern model were applied to the southern region and vice versa, the coefficients changed, leading to a decline in predictive skill. This difference in coefficients and predictive skill reinforces the necessity of using unique sets of predictors for each region instead of a singular model across all of Switzerland.*

Automatic predictor selection procedures (e.g. Recursive feature importance, LASSO, etc.) gave worse performing models than a manual stepwise approach combined with expert knowledge that was based on earlier considerations on optimal distribution separations of hail vs. no haildays (see Trefalt (2017)) and computed correlations (see Fig. A1, Fig. A2 and Table A4). Further discussions on the variable selection and their importance will follow in Sect. 6.~~, however, it is important to differentiate between the two domains. When applying the combination of variables of the model north on the region south and vice versa, the coefficients changed and the predictive skill declined. This again justifies the use of an individual combination of predictors for each region, rather than applying one model for the whole area of Switzerland~~

Line 330 is changed to: The selection of predictors followed the same procedure as in the logistic regression model.~~, however,~~ *Additionally,* for every variable that presented an effective degree of freedom (edf) > 1 a smoothing spline function was ~~used~~ *applied* to allow for non-linear effects.

Line 359 changed to: Observed values in thunderstorm environments often may exceed 1000Jkg−1.~~, however,~~ *In Europe,* severe hailstorms also occur with less *pronounced* CAPE (Taszarek et al., 2020a).

41. Table 1 caption: Avoid the construction with parentheses.
https://eos.org/opinions/parentheses-are-are-not-for-references-and-clarification-saving-space
Table 1 caption changed to: Coefficients, standard errors, *z*-values, and *p*-values of all covariates of the logistic regression model north and south. Positive~~ (negative)~~ signs indicate a positive ~~(negative)~~ relationship of the quantitative predictors with modelled hail occurrence *and vice versa.* ~~with hail occurrence relative to the reference category (April) for the categorical predictors.~~ [...]
All table captions will be adjusted accordingly.

42. Table 1: Italicize variables, such as z and v. Fix throughout. Adjusted throughout.

43. Table 2 caption: Why are bias, precision, and accuracy capitalized? Changed.

44. Lines 328, 331, and 344: There is no need for the apostrophe in GAMs, as it is not possessive. Find all occurrences in the manuscript and fix. Adjusted throughout.

45. Line 338: Be more specific. Most unstable CAPE determined over what depth?
Line 338 changed to: [...] CAPE is the most unstable convective available potential energy, c*omputed for parcels departing from model levels below the 350 hPa level.~~,~~[...]*

46. Line 340: d2m is not a sensible abbreviation for dewpoint temperature. I suggest T_{d,2m}, which would also be consistent with w_500. Moreover, the authors use the typical Td in Table A2. Be consistent internally within your manuscript and with common notation in the meteorological literature.
We will adjust the abbreviations for d2m in the revised manuscript.

47. Line 347: Why is “highly” in parentheses? Either delete the word or delete the parentheses. We deleted the word (“highly”) in line 347, as it does not make a difference if one predictor is highly significant or just significant.

48. Lines 347 and 571: Do you mean “variance” instead of “deviance”? Fix throughout the manuscript, if so. Indeed, we are talking about variance and not deviance. Changed throughout.

49. Lines 349–350: Why is GAM being redefined again? Definition removed.

50. Figure 3 and 4 need individual panel letters, and these figure numbers and letters need to be referred to in the text for the ease of the understanding of the reader. The panel letters will be added to the revised manuscript, together with their references in the text. We thank the reviewer for this suggestion.

51. Line 352: Italicize the y. Adjusted.

52. Line 353: Change “trough” to “through”. Adjusted.

53. Line 354: There are long sections of text without paragraph breaks. This paragraph is just one example that needs fixed. But, this problem exists elsewhere, too. We will restructure the mentioned paragraphs in the revised manuscript.

54. Line 361: What curve is being referred to? There are no figure and panel-letter citations in the sentences as far as I can see. Each reference to an interpretation of a figure needs to be cited so that the reader knows what they are looking at (e.g., Fig. 2a). Thank you for this suggestion. We added panel letters to the figures and reference citations in the text.

55. Line 370: Change “layer” to “level”. Adjusted.

56. Line 371: This text repeats earlier text. It also indicates a profound lack of understanding of what scales vertical motion acts, as discussed previously. We refer to comment C3.

57. At this point, I am frustrated by the lack of proofreading, so I will not repeat comments that I have made before beyond line 372. The previous comments should be applied throughout the manuscript during extensive proofreading. We will do extensive proofreading before the resubmission of the manuscript. Thank you for all your comments so far.

58. Lines 388–389: What are the citations for this information? We will add citations.

59. Line 392: Delete “We also need to mention that”, and revise the sentence. The sentence in line 392 was removed following a comment (C8) from reviewer 2. We acknowledge the need to avoid phrases like "we need to mention" in the manuscript, and we will ensure that future revisions adhere to this guidance. Thank you for bringing this to our attention.

60. Section 4.2: I am finding the manuscript tedious, particularly in this section. Is there anything the authors can do to improve the readability of this manuscript? Options could include shortening the text, putting more content into appendices or supplemental files, and improving interpretation of the results. We agree with the reviewer on the length of the manuscript and will restructure long paragraphs to improve the readability. Additionally, we will put the Appendix tables in the supplementary material and in turn will put all tables about coefficients of the statistical models and probably Figures 11 and 12 in the appendix.

61. Lines 461–462: Are two decimal places needed? Adjusted to one decimal place.

62. Line 466: Why is “e” raised to a power rather than “10”? Adjusted.

63. Line 472: No period at the end of this sentence. Adjusted.

64. Line 473: Insert “the” after “Using”. Added.

65. Figure 10 caption: “Modelled” needs to be capitalized. “linse" is misspelled. Adjusted.

66. Figures 10, 11, and 12: Why the unusual color bar? There is an abrupt color transition between red and blue in the middle of the bar. Why? The color scheme should have a smooth transition. Moreover, one could ask why a color scheme is needed in the first place? Is the value of the point not sufficient to indicate its value? Why does the dot need a color associated with it? We used the color scheme to show which values are below and which above the mean. We agree, however, that this might be leading to a bias in the readers' interpretations of the result. We will remove the color from the plot in the revised manuscript.

67. Line 487: Delete “it is important to mention that”. It is unnecessary. Line 487, “It is important to mention that” is deleted. Thank you for the suggestion.

68. Lines 525–549 is one long paragraph. It needs to be broken up. In fact, given the number of times that I have flagged paragraphs as being too short or too long, the authors should revisit the rules of writing paragraphs.
https://warwick.ac.uk/fac/soc/al/globalpad
rip/openhouse/academicenglishskills/writing/paragraphing/
I also suggest reading Gopen and Swan (1990) “The Science of Scientific Writing” (https://www.americanscientist.org/blog/the-long-view/the-science-of-scientific-writing) to gain further insight into better structuring your writing, particularly the use of topic and stress positions in sentences and paragraphs. Please indicate in your response that you have read this article and have implemented its guidance in the revised manuscript.
Unfortunately, the link to the first website you gave is no longer working.
The seven structural principles mentioned in the article on the science of scientific writing will be taken into account when restructuring paragraphs of the manuscript. We will focus specifically on the placement of old and new information in the topic vs. the stress position.

69. Line 531: I don’t understand “shear replacement”. Please explain. In many statistical models predicting hail occurrence, kinematic information is often represented by a wind shear predictor. In our models, wind shear was not chosen. Instead, the variable v_500 was the only kinematic information included in the southern model, making it a possible "replacement" candidate for wind shear. We agree that this terminology is misleading and have revised line 531 for clarity.

70. Line 538: Markowski and Richardson (2010) is a book. As they are not presenting original results, it would be inappropriate to write that they “showed” something. We agree. Line 538 is changed to: In fact, Markowski and Richardson (2010) and Dennis and Kumjian (2017) ~~showed~~ *explained* that wind shear primarily drives storm type and hail diameter.

71. Moreover, when citing a book, you should also cite the specific page number to which you are referring to. Otherwise, the reader will not be able to source your information from the entire book easily. We added page numbers to the reference.

72. Line 539: When you say “significant fraction”, please say what the specific number is. Houze et al. (1993) primarily focused on the directional movement of hailstorms in Switzerland, identifying 13 left-movers and 5 right-movers with true-hook structures out of 42 analyzed hailstorms. Since their study did not explicitly categorize hailstorms into supercellular versus ordinary or intermediate types, we

agree that it is more appropriate to refer to a more recent and relevant study for specific fractions. Therefore, we have revised line 539 as follows:
Feldmann et al. (2023) found that only ten percent of detected severe hailstorms in Switzerland are supercell type storms.

73. Lines 534–543: This text is poorly organized. It is not a systematic description of the literature that advances your argument. We agree. Section 6.1 on the interpretation of the lack of windshear in the model will be rewritten in the revised manuscript.

74. Line 544: What is "Alpine pumping"? Provide a citation and explain the result to nonspecialists. Or, delete. We added a citation and changed lines 544ff to: Lastly, with such complex terrain, shear might be driven by local conditions, *such as Alpine pumping* ~~(e.g. Alpine pumping)~~, that are too small to be resolved by ERA-5´s resolution. *Alpine pumping, a type of mountain-plain circulation, stems from differential heating and cooling of air masses between valleys and plains, creating horizontal pressure gradients. These gradients drive daytime winds from plains to valleys and nighttime winds in the opposite direction (Lugauer and Winkler 2004). Such local wind systems could be better resolved with high-resolution convection permitting models (with approx. 1x1km spatial resolution) than ERA5 data.* ~~This is something that can be better resolved with high-resolution convection permitting models (with approx. 1x1km spatial resolution).~~

75. Line 545: What is "orographic enhancement" (note correct term), and how does it create higher shear? Provide a citation and explain this mechanism to nonspecialists. After discussion with the coauthors, the sentence in line 545 will be removed.

76. Line 562: This is the first occurrence of "5D". It is not defined, and the authors use it twice on this page. Define and use throughout, or delete. We deleted the "5D" throughout, as the term is not important.

77. Line 564: Can you please provide an argument for this assumption that "models learn how to best deal with limitations of the datasets"? We refer to comment C4.

78. Lines 564–568: This text seems like just a list of random thoughts put here, without context, without evidence and without organization. We refer to the modifications implemented in response to comment C4.

79. Line 575: Why is the BI only defined here, even though it has been used throughout the manuscript? Define upon first usage. Definitions added to the manuscript where needed.

80. Line 590: Why is b before a? Put citations in alphabetical order. Adjusted.

81. Lines 610–611: Why not provide what the increase is? Can you normalize all the different studies to a percentage increase by year, as they all use different time periods? We added quantitative information on the increase in lines 610-611. Unfortunately, we are unable to normalize the different studies to a percentage increase per year due to the lack of complete data from the mentioned studies. We believe that performing such a normalization is more appropriate for a review paper.

82. Lines 618–619: Why is a more extensive period of observational data required? You are already using a decadal-length time series of data from different datasets. In response to comment C5, we have removed the sentence in line 618-619.

83. Line 661–662: The author should make this code freely available via a public repository. The authors decided to share the code on request, since the code only uses very commonly used packages from R to build and diagnose statistical models.

84. Line 669: AIC, BIC, and VIF are not defined here. Definitions added.

85. Line numbers are missing after line 665. Line numbers are added.

86. After equations for AIC and VIF: The comma should be on the same line as the equation. Adjusted.

87. Table A1 caption: Are TP, FP, FN, and TN needed? If not delete. If so, why not express the equations under "Explanation" in terms of these variables? The equations for performance metrics are calculated from contingency tables and typically use letters in the equations referring to TP, FP, FN and TN. It is important that the orientation of the contingency table is the same every time, otherwise the metrics are wrongly calculated, which is why using TP, FP, FN and TN leads to less confusion. We will express the metrics in terms of TP, FP, FN and TN instead.

88. Table A2: Why is deg01 used as an abbreviation for the height of the zero-degree isotherm? Why not use a more understandable expression, consistent with others? For example, z_{0°C}? The abbreviation for the zero degree level is not deg01 but deg0l ("l" as in level). All abbreviations come from the naming convention of the ERA5 files. We will change them to more standard abbreviations to make reading easier.

89. I am also similarly concerned about a number of other abbreviations that are not standard: FF, lfc (this and many others should be capitalized), mn2t, msl_mean, OMEGA (why all capitals?), etc. These inconsistencies should be made consistent, and standard abbreviations/variables used. Fix throughout the manuscript. We refer to the previous comment (88) and will adjust the necessary abbreviations.

90. If Taszarek is a coauthor, then he should not be listed in the acknowledgements as providing data, but in the Author Contributions in terms of data curation. Indeed, we thank the reviewer for finding this.

91. Also, Taszarek's data should be made available through the Code and data availability section.
https://www.natural-hazards-and-earth-system-sciences.net/policies/data_policy.html
Agreed, we will add references to thundeR data.