Authors replies to reviewer comments RC1

Review for EGUSPHERE-2023-2639

Characterizing hail-prone environments using convection-permitting reanalysis and overshooting top detections over south-central Europe

Antonio Giordani, Michael Kunz, Kristopher M. Bedka, Heinz Jürgen Punge, Tiziana Paccagnella, Valentina Pavan, Ines M. L. Cerenzia, and Silvana Di Sabatino

We want to thank the reviewer for his/her assessment of our manuscript. In the following we give our answers to the comments and recommendations that have been raised. Reviewer comments RC are **bold**, our replies AR are in *italic*.

General Comments:

RC: This study aims to develop a proxy for hail frequency, as determined by ESWD reports, based on high-resolution regional reanalysis atmospheric properties coincident with satellite OT detections. The authors demonstrate the proxy's improved capability in matching with hail reports relative to previous similar studies performed with coarser global reanalysis, although, by their admission, determined with focus on qualitative aspects. The thorough explanation of SPHERA variables and description of robust quality analysis are appreciated.

The concept of the filter could be introduced earlier or more gradually – it felt as though we were suddenly discussing the filter as a given before it was properly explained. That said, it is not convincing that the goal of obtaining a new proxy for hail has been satisfied quantitively. That is, the filter approach may be overly conservative compared to a random forest or other machine learning approach. It is perhaps good at removing cases where hail is unlikely, but not capable of skillful positive predictions. This task of separating events by their hailstone sizes through environmental analysis seems better suited for perhaps a logistic regression or a Bayesian analysis approach – as the layering of independently determined statistical filters seems to be too broad of a solution. Furthermore, adding additional years seems necessary given the high variability of year-to-year hail reporting, as a means to alleviate overfitting concerns, and to help the environmental stratifications.

I recommend major revision with consideration of the specific comments listed below. Furthermore, the paper can benefit from grammatical revision – some examples of which are listed as technical corrections.

AR: We thank the reviewer for the raised points. We agree that the concept of the filter in the original formulation of the manuscript seemed (at least partly) assumed as known to the reader, hence in the revised manuscript we clearly defined what is considered for "filter" since the Introduction. Additionally, we specified more clearly and unambiguously the aim of the proposed analysis, which is mainly the description of a new methodological approach tailored to exclude non-hail events, while is not foreseen for positive hail predictions nor for the definition of a climatology or long-term statistics for potential hail events, but can potentially be used for the purpose if adequately extended over multi-decadal time periods. In the revised manuscript we highlighted more clearly the objectives at the end of the Introduction section.
With respect to the conservative nature of the filter approach developed, we would like to point out

that the machine-learning alternatives mentioned by the reviewer might be more appealing in detecting true positive hail occasions in some situations and would certainly be interesting and valuable in the investigation of hail events as demonstrated by recent published literature in this context (e.g., Gagne et al., 2017; Gensini et al., 2021; Torralba et al., 2022; Scarino et al., 2023). A potential proposal for assessing the value of such an approach in this context could be to train a neural network with a large set of different environmental parameters (similarly e.g. to Prein & Holland, 2018), which however we believe is beyond the scope of the present paper, besides it would require an intensive effort. Furthermore, the necessary division into training and testing subsamples that a machine learning analysis would require is not possible for our case due to the short temporal series considered of only five years. Indeed, we are aware of the short time period of the analysis for which including additional years could be beneficial as pointed out by the reviewer. However, the extension of the temporal domain into the future (2021, 2022, 2023,..) is not feasible owing to SPHERA reanalysis temporal constraints (which ranges up to 2020 and for the moment future extensions of the dataset are not envisaged). On the other hand, extending into the past (2015, 2014, 2013...) we believe would not add significant benefits for the aim of the paper owing to the lower number of ESWD reports in the past years, as well as the systematical missing information on maximum hail size for the majority of reports (check e.g. Figure 1b in Torralba et al., 2023, who also show that, over Italy, only in 2019 and 2020 the number of hail reports amounts to more than 50% than those reported in the 20-year period 2000-2020).

In the revised manuscript we expanded the possibile future outlook in the Conclusion section, proposing that potential subsequent analyses could include the investigation of hail-related conditions through machine learning approches given the promising results obtained by documented recent literature that has been properly cited.

Additionally, we thank the reviewer for flagging many grammatical errors, in the revised manuscript we tried our best to improve the grammatical correctness of the manuscript, and we hope that it gained the desired correctness and clarity.

Specific Comments:

RC: Lines 101-108: The method for coupling satellite, SPHERA, and hail reports should be previewed somewhere around here in the introduction. "optimally tuning the combination" is mentioned, but in what kind of approach? Furthermore, "optimally tuning" is stated, but that seems to extend as far as layering 5 statistical thresholds, which does not suggest *tuning*.

AR: We agree that the term "optimally tuning" could probably be misleading and not completely correct in this context, hence in the revised manuscript we removed the expression and reformulated the highlighted sentence as suggested. Further, we introduced the coupling method and the filter concept more gradually.

RC: Fig 1/Line 137/Line 371: 1 h is said to be the cutoff for "high temporal accuracy," but what is the justification for this designation, as it seems like a rather large uncertainty, especially compared to the significantly greater sampling at 15-min temporal accuracy? What is the expectation for how 15-min vs. 1-h temporal accuracy would impact the hail environment characterizations?

AR: 1h was considered the threshold for ESWD temporal accuracy to be congruent with the spatio-temporal window (± 1h, 25 km) considered in the spatio-temporal matching criterion with OTs

presented in Section 4.2. It must be remembered that overshooting cloud tops observed near the tropopause (indicating strong updrafts and potentially associated with the beginning of hail growth) and hail reports observed at the surface are necessarily characterized by a certain spatio-temporal shift which implies a certain relaxation when matching the occurrence of one phenomenon with the other. However, we agree with the reviewer that for hail observation purposes 1h constitute a large temporal uncertainty, and the choice to consider the ESWD subsample with temporal accuracy ≤15 minutes, which matches also the sampling frequency of the satellite instrument detecting the OTs, could be more appropriate. In any case, it must be pointed out that the largest fraction of reports including information on their temporal accuracy is populated by reports with temporal accuracy ≤15 minutes (2077 reports: 77%), while 575 reports (21%) have temporal accuracy = 30min (421: 15%) or = 1 hour (154: 6%). For consistency, in the revised manuscript we adjusted the highlighted sentence accordingly and considered the ESWD subsample having temp. acc. ≤15 minutes when referring to "high temporal accuracy". Further, we considered both the ESWD subsamples having temporal accuracy ≤1h and those having temporal accuracy ≤15 min separately in the discussion of the statistical evaluation between OTs and hail reports matching.

RC: Line 139: "A substantial increase is noted for 2019 and, to a slightly lesser extent, for 2020." – Is this natural, or owed to ESWD hail reporting reasons? Showing more years may reveal more about the expected year-to-year variability of reports.

AR: The large increase in hail reporting in 2019 is believed to not be related to natural reasons but to a substantial increase in the public awareness and cooperation among European institutes collecting reports as observed in Púčik et al., 2019 (Section 3.a), and more recently by the ESWD reports statistics published by ESSL demonstrating the exponential increase in the number of hail reports in more recent years (2021, 2022, 2023): https://www.essl.org/cms/hailstorms-of-2023/ *(Figure reported below). Further, this statistics offers also a hint of the low impact that extending the temporal period of the analysis backwards could have (as discussed in the reply to the general comment). On the other hand, the difference between 2019 and 2020 is most likely imputable to natural year-to-year variability in hail occurrence over the region. Indeed, 2020 registered the lowest number of OTs (see Table below), which suggests less thunderstorm activity and hence that hail-prone conditions were more unlikely in the study region. In the revised manuscript we included an additional statement to clarify this point.*
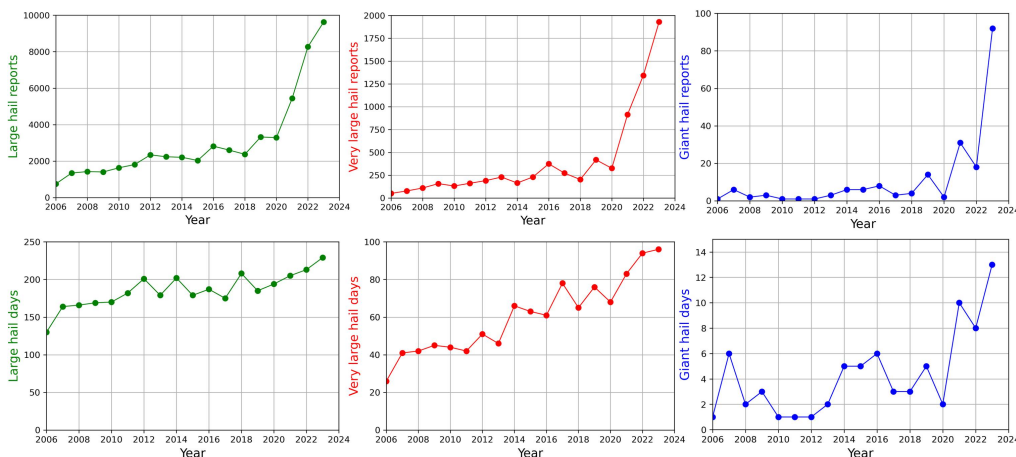
Year	2016	2017	2018	2019	2020
N of OTs	171.947	189.613	271.581	208.859	149.042

RC: Lines 193, 200, 237, 238: "… is included in the filter" – Again, reference to "the filter" and "full filter" is presented as a given without establishing what exactly is meant by the word. Reading further into the manuscript helps fill in and confirm the reader's intuition, but it would be helpful if there was a more explicit introduction to the filter concept – perhaps as a paragraph at the start of Section 3.

AR: The text has been revised accordingly (see the reply to the general comment and to the first specific comment).

RC: Line 257: Please explain what is meant by "the limitation imposed by choice of the H_0 threshold."

*AR: In the revised manuscript the specified sentence has been reformulated to avoid ambiguity and enhance clarity.*

RC: Lines 264-266: It isn't very convincing that the regions highlighted as having lowest OT removal are distinct from "noise". That is, the Gulf of Taranto and areas southwest of Sicily (among others) look comparably shaded to the ones the authors mention. Perhaps a discretized color bar in increments of 20% would help, but otherwise these discussion points should be removed or reframed in a way that's better suited to the image.

AR: We thank the reviewer for the raised point and we appreciate the suggestion. Figure 4 has been re-plotted accordingly (with discretized colobrar, reported below) and substituted in the text. Anyhow we agree with the reviewer: even with 5 discretized levels it is not evident that areas with lowest OT removal mentioned in the text (Po valley, Adriatic and Thyrrenean coastlines) are distinct from "noise" and are similar to other areas mentioned as southwest Sicily or the Gulf of Taranto. Hence, we reformulated the discussion in a more cautious fashion.
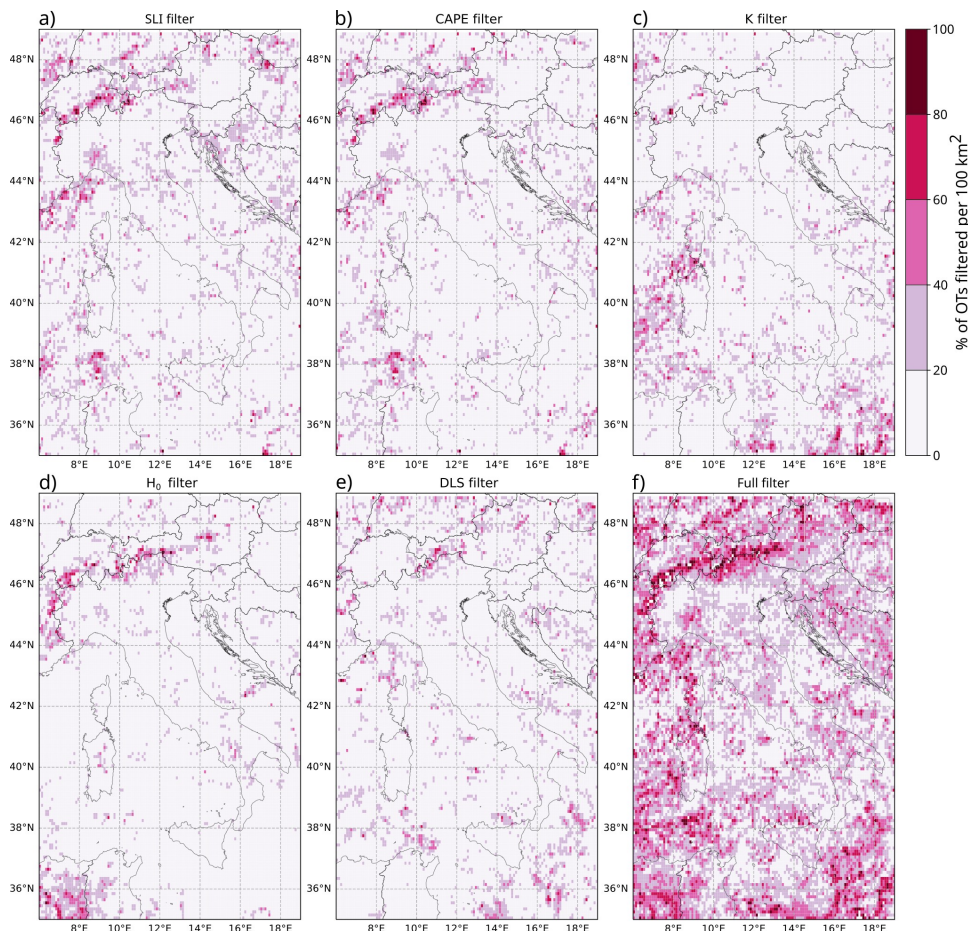
RC: Lines 271-272: There should be discussion of how the number of OT detections change over these same times of day and months, and how that may influence the interpretation of these fractions.

AR: Figure 5 has been expanded by adding two additional plots showing the total number of OTs per hour and month (reported below), and an additional relative discussion has been included in the text.
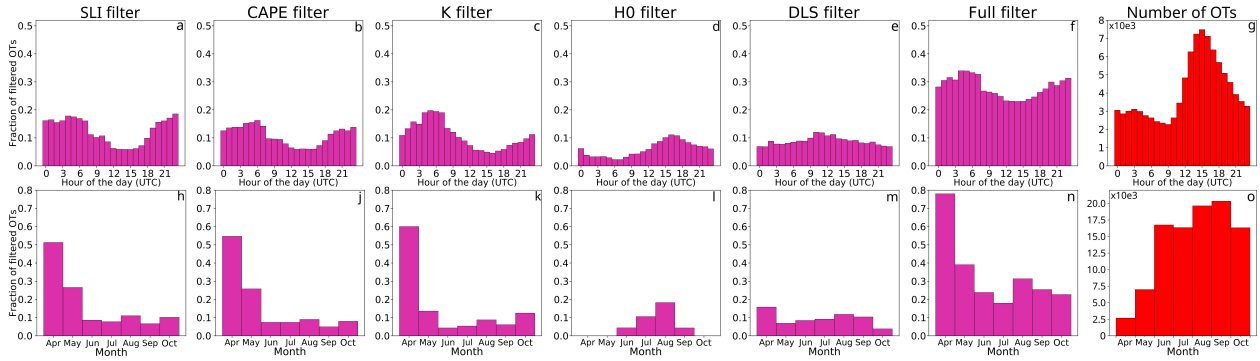
RC: Lines 360-363: It's not seen how "full independence of the two datasets" equates to no concern for overfitting. The authors are relating observed hail reports to coincident environmental analyses, that have spatial dependency, and then evaluating spatially dependent, environmental-analysis-filtered storm signatures. So it isn't convincing that "full independence" applies. The claim should be tested by doing the method but with one evaluation year held out from consideration for the filter. Perhaps even doing that for each year and evaluating the average performance. Adding additional years would also help in addressing this concern.

AR: Given the conservative nature of the proxy developed, which mainly focuses on filtering out the occasions where hail is unlikely, we believe that potential overfitting issues related to the usage of the same sample of ESWD hail reports data to define the environmental thresholds of the reanalysis convective parameters and to verifiy environmentally-filtered independent satellite cloud-top detections are unlikely to be dominant. Additionally, it is believed that operating a cross-validation by re-calculating the environmental thresholds leaving out single years/months from the sample analysis would be a considerably time-consuming task that would not change substantially the results owing to the small influence that slightly modifying the 5th or 95th percentiles (defining the thresholds) of the resulting environmental parameters distribution might have. As a kind of hint, we considered in the following the values of the environmental thresholds calculated by aggregating over the years 2016-2018 and 2016-2020 (considered in the manuscript) separately, which choice hardly have an impact in the resulting OT filtering and hence impact on the final proxy for hail:

| Parameter	2016-2018	2016-2020
CAPE_MU	618.42 J/kg	666.16 J/kg
LI	-2.08 °C	-2.16 °C
Kindex	32.5 °C	32.5 °C
DLS	9.25 m/s	7.56 m/s
H_0	4039 m	4098

However, we agree with the reviewer that "full independence" might be too hard of a statement, hence we revised the text accordingly being more cautious.

With respect to the addition into the analysis of additional years we believe it is out of the scope of the present work (refer to the reply to the general comment above). Further, we would like to point out that we explicitly proposed the present analysis only for testing the new hail proxy method (see the Introduction, line 104) rather than for developing a sound and robust climatology or statistics for potential hail which is not possible for 5-year data (see the Conclusion, lines 527-530).

RC: Table 2: Was a similar Table for day vs. night and/or month-to-month performance considered?

AR: We thank the reviewer for the interesting suggestion. We agree that reporting the inter-monthly and inter-daily variability of the performance between hail-filtered OTs and ESWD reports could be of interest to the reader. For this reason we included a new Figure (reported below) in the new Appendix B reporting how the hit rates change separating month-by-month and between day and night (when considering the spatio-temporal window for matching of 25 km and ± 1h), and included a brief discussion in Section 4.2.
Furthermore, as a consequence of this analysis, we thought of re-defining the hour ranges included in "daytime" and "nighttime": in order to try to isolate more appropriately the peak of convective activity over land detected in Figure 10, the daytime temporal range has been shifted to 10-21:45 UTC, and consequently nighttime to 22-9:45 UTC. Figure 9 has been revised accordingly.
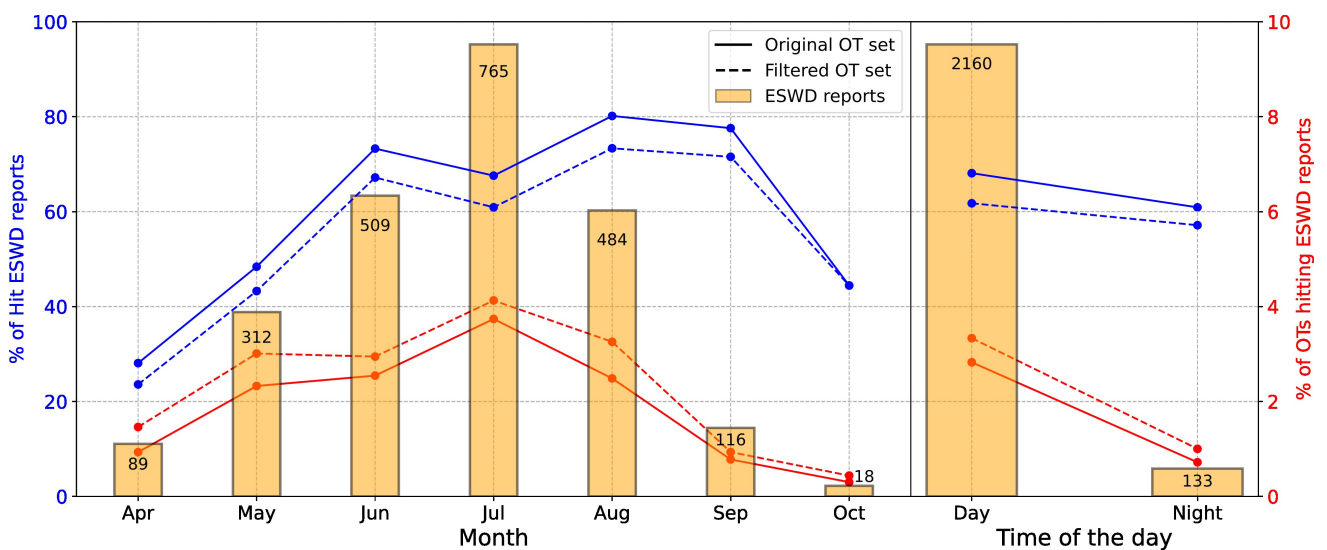
RC: Line 416: Please explain how it is decided which report with duplicate environmental conditions is kept.

AR: In case of multiple hail reports issued for the same storm and presenting the same environmental conditions within the spatio-temporal neighborhood considered for its ambient characterization, the report presenting the largest maximum hail diameter is kept. We are aware that a more detailed and precise information could benefit the environmental characterization, such as that possible by knowing the entire hail size spectrum for any issued hail report, but as of today is not feasible and the best possible information relies on the knowledge of the maximum hailstone size detected, which justifies the adopted choice. A note has been added in the text to explicitly account for this choice as suggested by the reviewer.

RC: Line 461: Please justify why "as expected."

AR: An appropriate reference has been added (Scarino et al., 2023).

RC: Lines 487-488: How can this conclusion be offered when the assessment used OTs to characterize the ΔT in missed cases? That is, if the OT are there and measured to be systematically warmer for missed cases, how can it be said that prominent OTs were not detected to characterize the events? This conclusion should be more clearly explained or removed. Furthermore, the preceding sentence requires grammatical revision, specifically "… are failed to be identified."

AR: The text has been revised accordingly, the highlighted sentece has been removed and the grammar of the preceding sentence adjusted.

Technical Corrections:

RC: Line 24: Though understood, the phrase "… and so its water holding capacity;" reads awkwardly in this context.

RC: Line 33: The phrase "… makes its observation still a major challenge…" probably can drop the word "*still*."

RC: Line 34: Similarly, the sentence with "… hail observing system is still missing" could benefit from a revision that does not include the word "*still*."

RC: Lines 35 and 86: Change "including" to "that include", otherwise it reads awkwardly.

RC: Line 43: The phrase "… constitute a precious way…" is distracting. Consider rephrasing and without use of the word "*precious*".

RC: Line 73: "These are…" – What are? No specific parameter was mentioned yet. Again, this may just be a structure issue. Consider revising this sentence and the one preceding.

RC: Line 88: Sentence requires grammatical revision.

RC: Line 94 and 98: What does "description" mean in these contexts? As in "characterization" or "depiction"?

RC: Lines 146 and 454: Suggest revision of the phrase, "…actually occurred OTs…", e.g., "true OTs".

RC: Line 97: Consortium for Small-Scale Modelling?

RC: Line 185: "Their different formulations… distinct parts of the numerical model equations…" – perhaps, though arguable. The sentence isn't necessary – consider removing.

RC: Line 223: It is unclear if "latter study" is referring to the approach of Punge et al. – check use of "latter" in this context.

RC: Line 246: Perhaps "… filter mainly over *certain areas* of the sea…", because largely the pattern across sea to land is rather continuous, especially for a) and b).

RC: Lines 284 and 524: "Anyhow" should usually be avoided in formal writing. Furthermore, the structure of the Line 284 sentence, with too many qualifiers followed by an "empty *this*," makes it difficult to follow. Please revise.

RC: Lines 465-466: The ending of this sentence does not read easily. Consider revising.

RC: Line 470: Conclusions are typically separate from Discussion.

AR: We thank the reviewer for the technical corrections and suggestions, all the proposed modifications have been accepted and the text revised accordingly.

AR: References

Gagne, D. J., McGovern, A., Haupt, S. E., Sobash, R. A., Williams, J. K., & Xue, M. (2017). Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. Weather and forecasting, 32(5), 1819-1840.

Gensini, V. A., Converse, C., Ashley, W. S., & Taszarek, M. (2021). Machine learning classification of significant tornadoes and hail in the United States using ERA5 proximity soundings. Weather and Forecasting, 36(6), 2143-2160.

Prein, A. F., & Holland, G. J. (2018). Global estimates of damaging hail hazard. Weather and Climate Extremes, 22, 10-23.

Púčik, T., Castellano, C., Groenemeijer, P., Kühne, T., Rädler, A. T., Antonescu, B., & Faust, E. (2019). Large hail incidence and its economic and societal impacts across Europe. Monthly Weather Review, 147(11), 3901-3916.

Scarino, B., Itterly, K., Bedka, K., Homeyer, C. R., Allen, J., Bang, S., & Cecil, D. (2023). Deriving Severe Hail Likelihood from Satellite Observations and Model Reanalysis Parameters Using a Deep Neural Network. Artificial Intelligence for the Earth Systems, 2(4), 220042.

Torralba, V., Hénin, R., Cantelli, A., Scoccimarro, E., Materia, S., Manzato, A., & Gualdi, S. (2023). Modelling hail hazard over Italy with ERA5 large-scale variables. Weather and Climate Extremes, 39, 100535.