

# Response to Reviewers

We sincerely thank the Editor and both Reviewers for their careful reading of our manuscript and for their constructive and insightful comments. We have revised the manuscript thoroughly to address every point raised. Below we provide a point-by-point response to each comment, indicating exactly which sections, paragraphs, and equations of the revised manuscript (R1.tex) have been modified. Reviewers' comments are reported in italics and prefixed by "Reviewer comment"; our responses follow each comment, together with a short summary of the corresponding changes implemented in the manuscript.

## Response to Reviewer #1

We thank Reviewer #1 for the constructive remarks, which helped us significantly clarify the methodological scope of RHITA, the limitations of the chosen reference dataset, and the role of the algorithmic parameters. All points have been addressed in the revised manuscript.

### Comment #1

**Reviewer comment:** *The study does not consider droughts, which are a major category of natural disasters and are included in the EM-DAT database. The authors should either provide a clear justification for excluding droughts, or discuss the feasibility and limitations of incorporating drought detection within the RHITA framework.*

**Response:** We thank the reviewer for raising this important point. We clarify in the revised manuscript that the present implementation of RHITA was designed for hazards with relatively rapid spatio-temporal evolution that can be detected directly from daily meteorological fields. Droughts differ substantially from these hazards because they evolve on longer time scales and result from cumulative hydrological processes (soil moisture depletion, evapotranspiration anomalies) with strong memory effects, and their identification typically relies on dedicated indices (SPI, SPEI, soil-moisture-based anomalies) rather than on direct daily threshold exceedances. The RHITA framework could in principle be extended to drought-related applications using accumulated or standardized indices over longer aggregation windows, but such an extension is outside the scope of the present work.

**Changes in manuscript:** A new dedicated paragraph has been added to the Introduction (Sec. 1, **LL 62-71**), explaining the methodological reasons why droughts are not included and citing van Loon (2015), Vicente-Serrano et al. (2010), McKee et al. (1993) and Cammalleri et al. (2023). A complementary sentence has been added near the end of the Conclusions (Sec. 6) listing the extension to slow-onset phenomena such as droughts as a natural future development of the framework. The word "droughts" was also removed from the list of hazards covered by the web tool in the preceding paragraph of Sec. 1 for consistency.

### Comment #2

**Reviewer comment:** *The Discussion section would benefit from a dedicated paragraph addressing uncertainties associated with ERA5 reanalysis and ECMWF forecasts, EM-DAT data biases, parameter selection in the algorithm. Additionally, the potential impact of these uncertainties on the results and conclusions should be explicitly discussed.*

**Response:** We agree and have substantially expanded the Discussion to address the three families of uncertainties (meteorological products, reference disaster database, algorithmic choices) and their impact on the results and conclusions.

**Changes in manuscript:** A new subsection "Sources of uncertainty and methodological caveats" (Sec. 5.1, Ssec:uncertainty, **LL 319-364**) has been added to the Discussion. It is structured in four paragraphs covering, respectively: (i) the limitations of the 0.25 deg daily ERA5 and ECMWF products for representing localized and short-lived convective phenomena, with explicit reference to known ERA5 biases over orographically complex regions and to forecast error growth; (ii) the limitations associated with EM-DAT as a socio-economically filtered archive and the mitigation strategies adopted (restriction to 2000-2023 and non-penalization of false positives); (iii) the methodological dependence of the detected events on the threshold  $q$ , the spatial grouping distance  $d$ , the minimum area  $A_{\min}$ , the centroid linking distance  $d_c$ , and the minimum-duration constraint, and the consequent caveats on the interpretation of continental-scale trends; (iv) the fact that RHITA characterizes the hazard component only and must be combined with exposure and vulnerability information for impact assessment. The Conclusions (Sec. 6) have been updated to explicitly refer to this new subsection.

### **Comment #3**

**Reviewer comment:** *Please clarify whether any preprocessing steps were applied to ECMWF forecasts and ERA5 data.*

**Response:** We have clarified this point in Section 2. The revised manuscript explicitly states how daily variables are constructed from the ECMWF operational forecasts (averaging of the eight 3-hourly temperature fields, computation of daily maximum wind speed from  $u$  and  $v$  components, accumulated precipitation taken from the 24 h lead-time field) and that ERA5 daily aggregates are retrieved directly with the CDSupdate Python package. No additional statistical bias correction, downscaling, or spatial filtering is applied prior to the detection step.

**Changes in manuscript:** Section 2.1 (Climate data, **LL 88-107**) has been revised: the paragraph describing the ECMWF forecasts has been extended with an explicit "The only preprocessing applied to the raw fields consists of standard daily aggregation" statement and a closing sentence ruling out additional bias correction; the ERA5 paragraph has been extended in the same way, with an explicit reference to the CDSupdate package URL and a closing sentence stating that all preprocessing reduces to standard daily aggregation of the native ERA5 fields.

### **Comment #4**

**Reviewer comment:** *The current evaluation relies primarily on sensitivity. Including additional performance metrics such as precision and F1 score would provide a more balanced and comprehensive assessment of the algorithm's performance.*

**Response:** We appreciate this suggestion and have clarified the rationale behind our evaluation strategy in the revised manuscript. Because EM-DAT cannot be regarded as a complete physical catalogue of meteorological extremes (many algorithm-detected events that do not appear in EM-DAT may correspond to genuine physical extremes that simply did

not generate sufficiently documented societal impacts), penalizing such cases as false positives would calibrate RHITA against socio-economic reporting practices rather than against the physical occurrence of extremes. For this reason we deliberately prioritize sensitivity together with the splitting, merging, and counting penalties in the cost function S (Eq. 3), and we now state explicitly that the resulting validation metrics should be interpreted as a lower bound on the algorithmic skill from a meteorological perspective.

**Changes in manuscript:** Section 3.2 (Parameters optimization, LL 192-203): a new explanatory paragraph has been added immediately after the description of Eq. (3) (cost function), explicitly discussing the meaning of precision and F1 in the strict machine-learning sense (with a reference to Juba and Le, 2019), the asymmetry between EM-DAT and a physical catalogue of extremes, and the rationale for not penalizing false positives. This rationale is also restated in the new Sec. 5.1 (uncertainty subsection in the Discussion).

#### **Comment #5**

**Reviewer comment:** Lines 22 to 24. Please add the appropriate references to support the statements made in this section.

**Response:** Additional references have been included in the revised manuscript to support the statements at the very beginning of the Introduction concerning the threats posed by extreme events, the 2024 disaster figures, and the hazard-exposure-vulnerability framework.

**Changes in manuscript:** Section 1 (Introduction, LL 14-24), first and second paragraphs: additional citations have been added to support these statements. Specifically: (i) the opening sentence on threats posed by extreme weather events now also cites Forzieri et al. (2017) and Ciscar et al. (2018); (ii) the 2024 disaster figures sentence now additionally cites the CRED 2024 review; (iii) the disaster-risk equation sentence now also cites Cardona et al. (2012); (iv) the sentence on climate change and exposure dynamics now also cites Seneviratne et al. (2021); (v) the sentence on increasing disaster complexity now additionally cites Pescaroli and Alexander (2018).

#### **Comment #6**

**Reviewer comment:** Line 66. It is recommended to include the date of data access/download for all datasets to ensure reproducibility and traceability.

**Response:** We agree. Access dates have been added systematically for all datasets.

**Changes in manuscript:** Section 2.1 (Climate data): the ECMWF open forecast paragraph now reports the last access date (31-01-2026) for the Open Data portal (L 90) in addition to the existing access date for the license URL; the ERA5 paragraph now reports the last access date (31-01-2026) for both the Copernicus Climate Data Store and the CDSupdate package URL (L 102). Section 2.2 (EM-DAT): the access date for the EM-DAT public portal (31-01-2026) has been added next to the URL in the opening sentence (L 111).

#### **Comment #7**

**Reviewer comment:** Line 110. Consider including a sensitivity analysis to evaluate how the results vary with different threshold choices. This would strengthen confidence in the robustness of the methodology.

**Response:** We thank the reviewer for this suggestion. Rather than introducing substantial additional analyses, we have clarified the rationale behind the chosen thresholds and tracking parameters and expanded the discussion regarding their influence on the detected climatology. The revised manuscript states explicitly that the identified events and associated trends are partly dependent on the selected thresholds, that RHITA is designed as a transparent and reproducible framework whose parameters can be re-tuned for different applications, and that the climatology should be regarded as one consistent realization within a family of algorithmic configurations.

**Changes in manuscript:** Section 3.1 (RHITA algorithm, LL 147-157), paragraph following the description of the four key parameters: the discussion has been extended to explicitly acknowledge sensitivity to the chosen reference period, to small variations of  $q$  around the adopted values, and to alternative configurations of  $d$ ,  $A_{\min}$ ,  $d_c$ . Section 5.1 (new uncertainty subsection) explicitly discusses the methodological uncertainty added by these choices on top of the statistical uncertainty of the Mann-Kendall and Theil-Sen estimators.

### **Comment #8**

**Reviewer comment:** Line 120. *The use of fixed quantile thresholds should be further justified.*

**Response:** We have expanded the justification for the use of the 0.99 and 0.01 quantiles in Section 3.1. The thresholds are now explicitly framed as a compromise between identifying sufficiently rare meteorological anomalies and retaining events whose societal relevance may arise from their duration or spatial extent rather than only from peak intensity. We also clarify that local quantile thresholds allow adaptation to the strong climatological contrasts across Europe and add references to previous studies using comparable percentile-based approaches.

**Changes in manuscript:** Section 3.1 (RHITA algorithm, LL 143-146), paragraph describing the threshold  $q$ : a new sentence has been added at the end of that paragraph explicitly justifying the choice of the 0.99 and 0.01 quantiles, citing previous percentile-based definitions of climate extremes (Russo et al. 2014; Perkins and Alexander 2013; Zhang et al. 2011; Schaer et al. 2016). The following paragraph (trade-off discussion) has been rewritten to highlight the role of local quantile thresholds in handling the strong climatological contrasts across Europe and to make explicit the dependence on the reference period and on parameter choices.

## **Response to Reviewer #2**

We thank Reviewer #2 for the detailed and balanced assessment of the manuscript and for the positive recognition of RHITA as an operational and open multi-hazard framework. We agree that several methodological limitations deserved clearer discussion and contextualization, and the revised manuscript has been modified extensively in this direction.

### **Comment #1**

**Reviewer comment:** *RHITA decides what an "event" is using chosen thresholds for extremeness, spatial grouping, minimum area, and temporal linking. Those choices are*

*sensible, but they are still subjective and they strongly affect what gets counted as one event versus many.*

**Response:** We fully agree. Event-based climatologies inevitably depend on methodological definitions, particularly regarding thresholds and spatio-temporal aggregation criteria. The revised manuscript states more clearly that RHITA provides a transparent and reproducible operational definition of hazards rather than an objective or unique definition of extreme events, and that the framework is intentionally designed to remain flexible and adaptable to different applications and hazard types beyond the scope of the web tool itself.

**Changes in manuscript:** Section 3.1 (RHITA algorithm): the paragraph (LL 147-157) discussing the trade-off underlying the choice of  $q$  has been rewritten to highlight that RHITA does not aim at a unique definition of extremes but at a transparent and reproducible operational framework, and that the resulting climatology is one consistent realization within a family of algorithmic choices. The new Sec. 5.1 (uncertainty subsection) contains a dedicated paragraph (LL 348-357) on the intrinsic dependence of the detected events on  $q$ ,  $d$ ,  $A_{\min}$ ,  $d_c$ , and the minimum-duration constraint. The Conclusions (Sec. 6, LL 367-370) have been updated accordingly: a sentence explicitly positioning RHITA as an operational/methodological contribution has been added in the first paragraph, and the second paragraph now restates that the climatology should be regarded as one consistent realization within a family of possible algorithmic configurations.

## **Comment #2**

**Reviewer comment:** *EM-DAT records disasters with societal impacts, but it has reporting gaps, inconsistencies, and country-level aggregation.*

**Response:** We agree and have strengthened this discussion throughout the manuscript. We now emphasize more explicitly that EM-DAT reflects reported disasters conditioned by societal exposure, vulnerability, and reporting practices, and that RHITA is calibrated against this impact-oriented database because the purpose of the optimization is to identify physically extreme hazard configurations associated with major societal events. The reasons for not penalizing false positives are now also stated explicitly.

**Changes in manuscript:** Section 3.2 (Parameters optimization): the paragraph following Eq. (3) (LL 192-203) on the non-penalization of false positives has been expanded to make this point explicit (see also response to Reviewer #1, comment #4). Section 5.1 (new uncertainty subsection, LL 338-347), second paragraph, contains a dedicated discussion of EM-DAT as a socio-economically filtered archive, of the reporting heterogeneities affecting it, and of the two mitigation strategies adopted (restriction to 2000-2023 and non-penalization of false positives).

## **Comment #3**

**Reviewer comment:** *Validation sensitivity for cold spells is much lower than for heatwaves or heavy precipitation.*

**Response:** We agree that this is an important limitation. The revised manuscript discusses this point more carefully and clarifies that the symmetric threshold choice was retained intentionally to preserve a homogeneous physical definition of warm and cold temperature

extremes. The implications for the interpretation of the cold spell statistics are now stated explicitly.

**Changes in manuscript:** Section 3.2 (Parameters optimization), paragraph discussing the joint optimization of heatwaves and cold spells: the discussion (**LL 218-223**) has been expanded to discuss why an asymmetric threshold (e.g., the 0.05 quantile for cold extremes) would mechanically increase recall but would also calibrate RHITA on socio-economic reporting practices that differ between northern and southern Europe and across seasons. A closing sentence has been added warning that the cold spell statistics should be interpreted with caution, particularly for regions where moderate winter conditions are routinely reported as disasters. The lower robustness of cold spell statistics is also explicitly mentioned in the new Sec. 5.1 (second paragraph) and in the limitations paragraph of the Conclusions (Sec. 6).

#### **Comment #4**

**Reviewer comment:** *For strong winds, they reuse the heavy precipitation parameter set because EM-DAT does not cleanly isolate wind-only disasters.*

**Response:** We agree and have expanded the discussion of this limitation. The revised manuscript states explicitly that the validation of wind events remains less robust because EM-DAT does not provide a clean separation between wind-driven and precipitation-driven storm disasters, and that the adopted parameter choice should be interpreted as a pragmatic approximation rather than as a hazard-specific calibration.

**Changes in manuscript:** Section 3.2 (Parameters optimization, **LL 224-231**), final paragraph on wind parameters: the text has been rewritten to make explicit that storm-related EM-DAT entries aggregate damages from both heavy precipitation and strong winds, that the wind configuration of RHITA is a pragmatic approximation assuming comparable synoptic-scale footprints, and that the implications are further discussed in the Discussion. The lower robustness of wind statistics is then reiterated in the new Sec. 5.1 (second paragraph) and in the limitations paragraph of the Conclusions (Sec. 6).

#### **Comment #5**

**Reviewer comment:** *The 0.25 deg data are good for broad, persistent systems, but not ideal for short-lived localized convective extremes.*

**Response:** We agree and have reinforced this point in both the Discussion and the Conclusions. The revised manuscript explicitly states that RHITA is expected to perform best for synoptic and large-scale persistent hazards and to be less reliable for short-lived, highly localized convective extremes.

**Changes in manuscript:** Section 5.1 (new uncertainty subsection, **LL 324-333**): explicitly discusses the consequences of the 0.25 deg resolution and daily aggregation for sub-grid convective phenomena, with references to ERA5 evaluation studies (Lavers et al. 2022; Betts et al. 2019). The Conclusions (Sec. 6), limitations paragraph (**LL 378-380**), has been rewritten to state explicitly that RHITA is expected to perform best for synoptic and large-scale persistent hazards and less reliably for highly localized convective extremes.

#### **Comment #6**

**Reviewer comment:** *The tool says how extreme the weather is, but not how exposed or vulnerable the affected population is.*

**Response:** We thank the reviewer for highlighting this important distinction. The revised manuscript clarifies more consistently that RHITA characterizes only the hazard component of disasters, and that disaster severity depends on the joint interaction of hazard, exposure, and vulnerability. RHITA does not aim to quantify impacts directly.

**Changes in manuscript:** Section 5.1 (new uncertainty subsection, **LL 358-364**): explicitly states that RHITA characterizes the hazard component only, that disaster severity reflects the joint contribution of hazard, exposure and vulnerability, and that the metrics provided by RHITA should be combined with exposure and vulnerability information when used to support risk assessment. The same point is reiterated in the limitations paragraph of the Conclusions (Sec. 6). The Introduction (Sec. 1) already framed this distinction but additional citations (Cardona et al. 2012; Seneviratne et al. 2021) have been added to reinforce it.

#### **Comment #7**

**Reviewer comment:** *Some of the Europe-wide trend analyses may conceal important regional or seasonal differences.*

**Response:** We agree and have clarified this point in the revised Discussion. We now stress explicitly that continental-scale aggregation can mask important regional and seasonal heterogeneity, particularly for precipitation- and wind-related hazards, and that the absence of robust European-scale trends should not be interpreted as evidence of the absence of regional changes.

**Changes in manuscript:** Section 5 (Discussion, **LL 310-318**): a new paragraph has been added immediately before the new Sec. 5.1, explicitly stating that the absence of significant continental-scale trends should not be interpreted as evidence of the absence of regional or seasonal changes, that aggregation across Mediterranean, Atlantic, continental and northern European climates and across cool and warm seasons can mask substantial heterogeneity, and noting that the regional analysis in Sec. 4.2 already shows that meaningful signals do emerge once spatial aggregation is reduced. A finer seasonal/regional decomposition is mentioned as a natural direction for follow-up studies based on the open RHITA archive. The Conclusions (Sec. 6) have also been updated to include "seasonal stratification" in the list of factors to which trend detection is sensitive.

#### **Comment #8**

**Reviewer comment:** *Its big contribution is building an operational framework and web tool. The novelty is more methodological than conceptual.*

**Response:** We agree with this assessment. The revised manuscript positions RHITA more clearly as an operational and methodological contribution aimed at unifying hazard detection, three-dimensional tracking, climatological characterization, and near real-time visualization within a common open framework.

**Changes in manuscript:** Section 1 (Introduction **LL 43-45**): a new sentence has been added at the end of the paragraph introducing RHITA explicitly stating that "the main contribution of RHITA is methodological and operational: it unifies hazard detection, three-

dimensional tracking, climatological characterization and near real-time visualization within a single open framework, rather than introducing a conceptually new definition of extreme events." The Conclusions (Sec. 6) have been rewritten accordingly: the first paragraph now opens with a clear statement positioning RHITA as a primarily operational and methodological contribution that unifies detection, tracking, climatology, and near real-time visualization within a single open and reproducible workflow.