Response to Reviewers' Comments [Paper # essd-2026-607]

We sincerely thank the two reviewers for their constructive and insightful comments, which have greatly helped us improve the quality of the CFDID database and our manuscript. We also wish to express our sincere appreciation to the editors for their professionalism and efficiency in handling our manuscript. The main revisions are summarized below:

1. LLM-extraction evaluation and data quality of CFDID. We have added a quantitative field-level evaluation of the LLM outputs based on complete manual checking against the original source documents and clarified the database construction workflow. We have re-examined the database, and recomputed the associated descriptive statistics to ensure consistency between the released dataset and the manuscript. These revisions strengthen the reliability assessment of the extraction process and the documentation of data quality.

2. Event-level comparison with EM-DAT and the added value of CFDID. We have added a systematic event-level comparison between CFDID and EM-DAT based on manually checked record correspondences. The analysis distinguishes records with and without identified counterparts, compares record counts across impact-magnitude classes, and quantifies the associated reported impacts. The results show that CFDID provides substantial event-level impact information through records without identified EM-DAT counterparts. This comparison offers more direct evidence of CFDID's contribution to flood-impact research in China, beyond differences in annual totals and total record counts.

3. Supporting information, repository documentation, and reproducibility. We have supplemented the CFDID repository with official annual benchmark totals, a README, a machine-readable data dictionary, a validation script, version-history information and software-environment information. We have also clarified the international database versions and filtering rules.

We hope that these revisions address the reviewers' concerns and improve the quality of the manuscript and the database. We sincerely thank you again for your constructive comments and valuable suggestions. Detailed, point-by-point responses and the corresponding manuscript revisions are provided below.

Reviewer 1

R1C0:

Comment: *This is a valuable and carefully executed contribution. The manuscript is clearly written, the developed verification workflow is transparent and reproducible (including the extraction prompt in Appendix B), and the coverage assessment against official totals and against EM-DAT, DFO, and Wikimpacts gives readers a clear sense of the database's added value. The inclusion of agricultural impact indicators, rarely available in global databases, is a genuine strength. I think this paper just needs some minor revisions (see details below).*

Response:

We sincerely thank you for your favorable assessment of the manuscript and for recognizing the value of CFDID. Your careful comments have helped us improve the data context, quality assessment, and consistency of the manuscript. The principal revisions in response to your comments are summarized below:

1. Regarding the context and complementary data sources: We have introduced the *Global Drought and Flood Catalogue* in the Introduction and discussed its potential role in future cross-validation of event timing and spatial extent in Sect. 4.4 (R1C1).

2. Regarding data-quality assessment: We have added a quantitative assessment of the LLM outputs based on complete field-by-field manual checking, with the evaluation procedure and field-specific accuracies presented in Sect. 2.4.3 and Table 3. We have also clarified that the comparison with official annual totals is partly internal rather than fully independent validation, and discussed the potential and limitations of using external sources (R1C2).

3. Regarding database consistency and future updates: We have rechecked the impact-field counts and the classification of the Sichuan dam-break record, and revised the corresponding descriptions. We have also added version-history information to the Zenodo record to help users distinguish releases and track future additions and corrections (R1C3–R1C4).

We sincerely thank you again for your careful review and constructive suggestions. Detailed responses to each comment and the corresponding manuscript revisions are provided below.

R1C1:

Comment: *The introduction and Table 3 review the Emergency Events Database (EMDAT), the Dartmouth Flood Observatory (DFO), and Wikimpacts, but do not mention other global database, such as the Global Drought and Flood Catalogue (https://doi.org/10.1175/BAMS-D-18-0269.1). This data does not report socioeconomic impact indicators, so it would not slot directly into the Table 3 coverage comparison, but it is a natural addition to the landscape review alongside DFO (lines 37-43), since both offer an independent, physically derived hazard perspective at the global scale that complements the impact focused databases discussed here. The authors may wish to cite it and note it as a potential resource for future cross-validation of event timing and spatial extent.*

Response:

We sincerely thank you for this helpful suggestion. We agree that the *Global Drought and Flood Catalogue (GDFC)* is an important complement to existing global flood and drought data resources. Unlike impact-focused databases such as CFDID, EM-DAT, and Wikimpacts, GDFC provides physically derived records of hydrological extremes and does not include socioeconomic impact indicators. It therefore cannot be directly included in the quantitative impact comparison. Nevertheless, such physically derived hazard datasets provide an independent reference for cross-validating the timing and spatial extent of events recorded in impact databases.

Following your suggestion, we have added GDFC to the Introduction to provide a more complete overview of existing global flood-related datasets. We have also added a statement in Sect. 4.4, "Limitations and future improvements," noting that GDFC and similar physically derived hazard datasets could provide independent information for future cross-validation and enrichment of CFDID events.

1 Introduction: Lines 41-43:

"The Global Drought and Flood Catalogue provides long-term, physically derived records of hydrological extremes (He et al., 2020), but does not include socioeconomic impact information."

4.4 Limitations and future improvements: Lines 389-390:

"Physically derived hazard datasets, such as the Global Drought and Flood Catalogue (He et al., 2020), can also provide independent information on event timing and spatial extent."

R1C2:

Comment: *Validation against independent sources. The main validation exercise (Section 3.2, Section 4.1, Fig. 7) compares CFDID against official annual national totals and against EM-DAT. It would help readers to note explicitly that the official totals used as the benchmark come from the same authority families that supply the event-level records, so this comparison is partly internal. Have the authors considered validation against sources outside this reporting chain, such as remote-sensing-derived inundation extent, insurance or reinsurance loss data, or provincial-level records? A quantified accuracy assessment of the OCR-LLM extraction step itself (e.g., the share of extracted fields requiring manual correction, or precision/recall against a manually labeled subset) would also strengthen the data quality claims, which are currently described only qualitatively.*

Response:

We sincerely thank you for this constructive comment. We agree that both the annual totals and the event-level records used to construct CFDID originate from the governmental

reporting system, although they are compiled at different aggregation levels. In China, long-term nationwide disaster-impact statistics are primarily compiled and verified by government agencies; therefore, the official national totals represent the most systematic and authoritative benchmark currently available for evaluating the national-scale consistency of historical flood-impact records. We have clarified in Sect. 2.5 that this comparison is partly internal.

We considered independent sources outside this reporting chain. For remote-sensing-based information, Qiu et al. (2025) compared Sentinel-1-derived flooded cropland areas with officially reported affected cropland areas in China during 2017–2021 and reported good consistency between the two. This published comparison provides contextual support for the use of official affected-cropland statistics as a benchmark. We have added this study to Sect. 2.5 as independent supporting evidence for the affected-cropland indicator.

We did not use insurance or reinsurance losses as an independent national benchmark because we could not identify a long-term, nationally consistent, and openly accessible dataset suitable for systematic comparison with CFDID. Moreover, insured losses do not necessarily represent total economic losses, particularly where insurance coverage is incomplete.

Provincial-level official statistics were also not used as an independent validation source for two reasons. First, they generally belong to the same governmental reporting chain as the national totals and therefore would not provide a fully independent benchmark. Second, some L1 records in CFDID cover multiple provinces, while their reported impacts cannot always be disaggregated among the individual provinces. A direct comparison with provincial totals could therefore introduce inconsistencies in the spatial allocation of impacts.

In addition, following your suggestion, we conducted a complete quantitative evaluation of the LLM-based extraction. All 1,912 records generated at the extraction stage were manually checked field by field against the original source documents before data fusion. Across 18 extracted fields, 34,416 field instances were evaluated. The overall field-level extraction accuracy was 99.25%, with field-specific accuracies ranging from 97.23% for “Specific location” to 99.95%

for “Typhoon information”, “Damaged cropland area”, and “Failed cropland area”. These results have been added to Sect. 2.4.3 and summarized in the new Table 3.

2.4.3 Data extraction: Lines 167-175:

“To evaluate extraction accuracy, we manually checked the LLM-based outputs field by field against the source documents before data fusion. We evaluated all 18 fields across 1,912 records, covering 34,416 field instances in total. A field output was considered correct if it accurately reflected the source information, including a null value when the relevant information was not reported. The overall field-level extraction accuracy was 99.25%, with field-specific accuracies ranging from 97.23% to 99.95% (Table 3). The field “Specific location” had the lowest accuracy (97.23%). This mainly arose because some source documents reported both locations experiencing heavy rainfall and locations where disaster impacts occurred; the LLM occasionally extracted the former, whereas CFDID records the latter as the disaster location. These discrepancies were identified and corrected during manual verification.

Table 3. Field-level accuracy of LLM-based information extraction across all 1,912 records.”

Field	Accuracy (%)	Field	Accuracy (%)
Start year	98.64	Typhoon information	99.95
Start month	97.91	Deaths and missing persons	99.16
Start day	98.43	Affected population	99.79
End year	99.37	Affected cropland area	99.74
End month	99.22	Damaged cropland area	99.95
End day	99.11	Failed cropland area	99.95
Province	99.58	Collapsed dwellings	99.27
Specific location	97.23	Damaged dwellings	99.74
Disaster category	99.69	Direct economic losses	99.74

2.5 Assessment relative to official annual totals: Lines 216-220:

“The official annual totals used as benchmarks originate from the same governmental reporting system as the event-level records used to construct CFDID. Therefore, the comparison between CFDID and official annual totals is partly internal and should not be interpreted as a fully independent validation. Qiu et al. (2025) reported good consistency between Sentinel-1-derived

flooded cropland areas and officially reported affected cropland areas in China during 2017–2021, which offers independent evidence for the affected-cropland indicator."

R1C3:

Comment: *Two smaller items are worth double-checking for the authors.*

a. Section 3.1 states that 397 records contain failed cropland area and, separately, that 397 records contain damaged dwellings. These are unrelated indicators sharing an identical count. Please confirm this is not a copy-paste error.

b. Section 3.5 reports 120 total province-level records for Sichuan (Fig. 6a) but only 119 rainstorm flood records for Sichuan in the following paragraph, and Sichuan does not appear among the snowmelt, ice-jam, or dam-break totals elsewhere in the text. Please clarify which category accounts for the remaining record, or correct the figure if needed.

Response:

We sincerely thank you for the careful checking. We re-examined the database and the corresponding statistics.

a. The statement in the original manuscript that "397 records contain failed cropland area" was a typographical error. The correct number is 395, whereas the number of records containing damaged dwellings remains 397. We have corrected this value in Sect. 3.1.

b. Sichuan has 120 province-level records in total, of which 119 are rainstorm flood records and one is a dam-break flood record (Event ID: 2018-062). In the original manuscript, this dam-break flood record was mistakenly attributed to Qinghai rather than Sichuan in the text describing the spatial distribution of dam-break floods. We have corrected this error in Sect. 3.5.

3.1 Overview of CFDID: Lines 225-228:

"Among the 1716 records, 1205 records contain deaths and missing persons; 1419 records contain affected population; 1241 records contain affected cropland area; 87 records contain damaged

cropland area; 395 records contain failed cropland area; 969 records contain collapsed dwellings; 397 records contain damaged dwellings; and 1444 records contain direct economic losses."

3.5 Spatial distribution: Lines 301-303:

"Dam-break flood records are mainly found in western China, with three records in Xizang and one record each in Xinjiang, Yunnan, and Sichuan (Fig. 6f)."

R1C4:

Comment: *A very minor issue. Given that the repository will be continuously updated by year, a brief version-history note or changelog in the Zenodo record (beyond Table 2's field list) would help users track what changes between annual updates.*

Response:

We sincerely thank you for this helpful suggestion. We agree that a clear version history is important for a database that will be updated, as it allows users to track changes between releases. Following your suggestion, we have added a "Version history" note to the Zenodo dataset description.

Zenodo:

"**Version history:**

Version 1.0 (30 July 2026): Initial release of CFDID, covering 1994–2023, with 1,716 event records, including 1,391 L1 records, compiled from 66 official documents.

Version 1.1 (23 September 2026): Revised release accompanying manuscript revision. Corrected selected impact values and event dates, and added official annual benchmark totals, the CFDID–EM-DAT record correspondence table, a README, a machine-readable data dictionary, database validation code, software dependency information, and updated repository metadata. The total number of records remains unchanged."

Reviewer 2

General comments: *CFDID makes available, in open and structured form, event-scale flood impact information that until now existed only in printed Chinese official yearbooks — the complete annual runs of three national series for 1994–2023, for a country that is both heavily flood-affected and likely under-represented in existing international impact databases, notably because of entry criteria and language biases. The resulting resource can be very useful for disaster risk research, management, and planning. To my knowledge—which is quite limited regarding the Chinese disaster loss data landscape—no alternative is openly available for China, except perhaps for urban floods. It also provides indicators that are missing from global databases, such as the agricultural impact indicators.*

The construction workflow (OCR, LLM-based extraction against a fixed output schema, cross-source fusion, and manual verification) is a careful and well-documented application of components that can be considered state-of-the-art practices, rather than a methodological advance. The dataset shares some design choices with EM-DAT and Wikimpacts; Hence, the paper is primarily valued as a data contribution, as expected for a data paper, and the workflow description will still be instructive to others attempting similar transcription projects, though not sufficiently evaluated.

The deposited file is structurally sound but would benefit from a clear README to make the Zenodo repository more self-contained and less dependent on the paper. Event IDs are unique and well-formed; every L2 and L3 record resolves to a parent; no sub-event impact exceeds its parent; there are no negative values or out-of-range dates; the 31 province names are used consistently, and the constant-price column is internally consistent with the CPI adjustment described in Eq. (1). The reference list is likewise sound: every reference is cited, and every citation resolves. I congratulate the authors for the care and rigor they put into the checks. The paper reads really well, without any grammar or spelling issues identified.

While I endorse the paper for publication in ESSD and congratulate the authors on their work, I have 2 general comments that stand in the way of a direct recommendation. None of them require major changes to the dataset or its workflow; rather, they relate to information disclosure and improved comparisons. I therefore recommend a minor revision.

Response:

We sincerely thank you for your detailed and thoughtful review, your careful inspection of the deposited database, and your positive assessment of its potential value. We agree that the main contribution is the openly available, manually verified event-level database, rather than a new extraction algorithm. Your comments have helped us strengthen the evaluation, comparison, and documentation of this data contribution. The main revisions are summarized below:

1. Regarding extraction reliability: We have added a quantitative field-level assessment based on manual checking of all 1,912 extraction-stage records against the original source documents. The evaluation covers 34,416 field instances across 18 fields, including correctly identified missing values, and yields an overall field-level accuracy of 99.25%. We report the evaluation procedure, field-specific accuracies, and the main source of location-extraction errors in Sect. 2.4.3 and Table 3 (GC1).

2. Regarding database comparability and complementary information: We have clarified the international database versions, filtering rules, and assumptions concerning EM-DAT's numerical entry thresholds. We have expanded Sect. 4.2 and Fig. 7 to compare record counts by magnitude and identified correspondence between EM-DAT and CFDID, and quantified the reported impacts of records without counterparts in both databases relative to official national totals. We also examine correspondences among CFDID records meeting at least one numerical threshold and discuss the complementary information provided by the two databases (GC2 and SC3–SC5).

3. Regarding reproducibility and data reuse: We have rechecked the descriptive statistics and addressed the identified record-level issues. We have also supplemented the repository with

the official annual benchmark totals, a README, a machine-readable data dictionary, a validation script, and software dependencies (SC2, SC7, and TC1–TC6).

In addition, we have replaced the ambiguous "coverage ratio" terminology and added some prior LLM-based extraction studies (SC1 and SC6).

We sincerely thank you again for your valuable suggestions. Detailed responses to each comment and the corresponding revisions are provided below.

GC1:

Comment: *Lack of extraction evaluation.*

The manuscript provides no quantitative estimate of extraction accuracy for the DeepSeek model or inter-author agreement, while ESSD explicitly asks for error estimates and documented sources of uncertainty. I believe this could be reported easily based on the manual revisions (section 2.4.3) that were executed, to provide a benchmark for comparison with similar works. Since the workflow is offered as a methodological reference for others, its reliability is what readers need to reuse it.

Response:

We sincerely thank you for this constructive comment. We agree that a quantitative evaluation of the LLM-based extraction is important for assessing the reliability and reproducibility of the workflow.

We therefore conducted a complete field-level accuracy assessment based on the manual verification process. All 1,912 records generated at the extraction stage were manually checked field by field against the original source documents before data fusion. Across 18 extracted fields, 34,416 field instances were evaluated. The overall field-level extraction accuracy was 99.25%. Field-specific accuracies ranged from 97.23% for "Specific location" to 99.95% for "Typhoon

information", "Damaged cropland area", and "Failed cropland area". We have added the field-specific accuracies to Sect. 2.4.3 and summarized them in a new table.

The relatively lower accuracy for "Specific location" mainly reflects the complexity of spatial descriptions in the original documents. Some source texts mention both locations where heavy rainfall occurred and locations where disaster impacts were actually reported. The LLM occasionally extracted the former, whereas CFDID records the latter as the disaster location. These cases were identified and corrected during the manual verification process.

2.4.3 Data extraction: Lines 167-175:

"To evaluate extraction accuracy, we manually checked the LLM-based outputs field by field against the source documents before data fusion. We evaluated all 18 fields across 1,912 records, covering 34,416 field instances in total. A field output was considered correct if it accurately reflected the source information, including a null value when the relevant information was not reported. The overall field-level extraction accuracy was 99.25%, with field-specific accuracies ranging from 97.23% to 99.95% (Table 3). The field "Specific location" had the lowest accuracy (97.23%). This mainly arose because some source documents reported both locations experiencing heavy rainfall and locations where disaster impacts occurred; the LLM occasionally extracted the former, whereas CFDID records the latter as the disaster location. These discrepancies were identified and corrected during manual verification.

Table 3. Field-level accuracy of LLM-based information extraction across all 1,912 records.

Field	Accuracy (%)	Field	Accuracy (%)
Start year	98.64	Typhoon information	99.95
Start month	97.91	Deaths and missing persons	99.16
Start day	98.43	Affected population	99.79
End year	99.37	Affected cropland area	99.74
End month	99.22	Damaged cropland area	99.95
End day	99.11	Failed cropland area	99.95
Province	99.58	Collapsed dwellings	99.27
Specific location	97.23	Damaged dwellings	99.74
Disaster category	99.69	Direct economic losses	99.74

"

GC2:

Comment: *Lack of detail, fairness, and reproducibility for the comparison with other data sources.*

Versions, endpoints, and access dates for external datasets are not disclosed. Some figures could not be reproduced from the archived file, while the comparison with international databases rests on unstated assumptions — including which disaster types were selected, which database version was used, and whether the entry criteria were met. Moreover, the comparison relies on event counts, an indicator that is neither robust nor unbiased due to differences in reporting and aggregation practices. This GC maps to several SCs reported in fragments in my specific comments, providing more specific pathways for corrections.

Response:

We sincerely thank you for this constructive comment. We agree that the comparison with other databases requires explicit clarifications, more informative record-level analyses, and reproducible supporting materials. In response to this general concern and the related specific comments, we have made the following revisions:

1. Regarding comparison criteria and interpretation: We now specify the disaster categories and record levels selected from each international database, together with the versions and access information (SC3).

2. Regarding record-level complementarity: We have extended Sect. 4.2 beyond annual totals by examining manually checked correspondences between CFDID and EM-DAT records and comparing record counts across multiple impact-magnitude classes in Fig. 7d–f. We quantify the reported impacts of records without identified counterparts in both databases relative to the corresponding official national totals. We also examine CFDID records meeting at least one of EM-DAT's two numerical thresholds to assess whether complementary information remains evident after accounting for these thresholds. The matching procedure and the scope for further

comparison of impacts within matched record groups are described in Sect. 4.2 and Appendix C (SC3–SC5).

These revisions distinguish the assumptions governing the comparisons and the complementary information identified in the records. Further details are provided in our responses to the corresponding specific comments.

SC1:

Comment: *Terminology: coverage ratios.*

While Eqs. 2 and 3 make it explicit, I, in the first place, was not able to understand, in the abstract and at several places in the manuscript, what was meant by coverage ratios. First, the indicator is a percentage, not a ratio; second, "coverage" can ambiguously refer to whether an event is covered in a database (i.e., its occurrence) or to spatial coverage (e.g., L35 uses coverage to mean geographic coverage). Hence, I would suggest something more explicit and accurate, like "impact reporting percentages relative to China's official totals…".

Response:

We sincerely thank you for this constructive comment. We agree that the term "coverage ratio" used in the original manuscript could be ambiguous. First, the metric is expressed as a percentage rather than a ratio. Second, "coverage" may refer to whether events are included in a database or to spatial coverage, whereas the metric used here specifically represents the disaster impacts recorded in a database as a percentage of the corresponding official national totals.

We have therefore removed the term "coverage ratio" throughout the manuscript and replaced it with more explicit wording based on the percentage relative to the corresponding official totals. We have also renamed Sect. 2.5 as "Assessment relative to official annual totals". Accordingly, we revised the relevant text in the Abstract, Introduction, Sect. 2.5, Sect. 3.2, Fig. 2,

Table 4, and the Discussion to avoid ambiguity between event inclusion, spatial coverage, and the comparison of recorded impact magnitudes with official totals.

Abstract: Lines 18-22:

"Relative to the corresponding official national totals during 1994–2023, the L1 records in CFDID account for 68.49% of deaths and missing persons, 72.24% of affected population, 60.70% of affected cropland area, 65.67% of collapsed dwellings, and 74.86% of direct economic losses. Compared with commonly used international flood impact databases, CFDID contains more detailed event records and accounts for higher percentages of the corresponding official totals for most impact indicators."

1 Introduction: Lines 60-63:

"Third, using official annual flood impact totals as benchmarks, we compare the percentages of these totals represented in CFDID and several international databases, providing a quantitative basis for understanding the complementarity among different disaster impact databases."

2.5 Assessment relative to official annual totals: Lines 202-212:

"We calculated annual and total percentages relative to the corresponding official national totals to quantify the disaster impacts represented in CFDID and international disaster impact databases. Different databases use different event structures. For CFDID, only L1 records were used in the numerator, and nested L2 and L3 records were excluded to avoid double counting between parent events and their sub-events. For Wikimpacts, which also has a hierarchical structure, we used L2 national-level records for China. For EM-DAT and DFO, which do not have hierarchical event levels, the original event records were used directly. Percentages were calculated only for indicators available in each database. For database d, indicator k, and year t, the annual percentage relative to the corresponding official annual total is defined as:

$$P_t^{d,k} = \frac{\sum_{i=1}^{n_t} Impact_{i,t}^{d,k}}{Official_t^k} \times 100\% \quad (2)$$

The total percentage relative to the corresponding official totals over 1994–2023 is calculated as:

$$TP^{d,k} = \frac{\sum_{t=1994}^{2023} \sum_{i=1}^{n_t} Impact_{i,t}^{d,k}}{\sum_{t=1994}^{2023} Official_t^k} \times 100\% \quad (3)"$$

3.2 Percentages relative to official annual totals: Lines 235-249:

"The L1 records in CFDID represent substantial percentages of the official national totals for the five core impact indicators. Over 1994–2023, the corresponding total percentages are 68.49% for deaths and missing persons, 72.24% for affected population, 60.70% for affected cropland area, 65.67% for collapsed dwellings, and 74.86% for direct economic losses (Fig. 2a). Comparisons with international disaster impact databases are provided in Sect. 4.1.

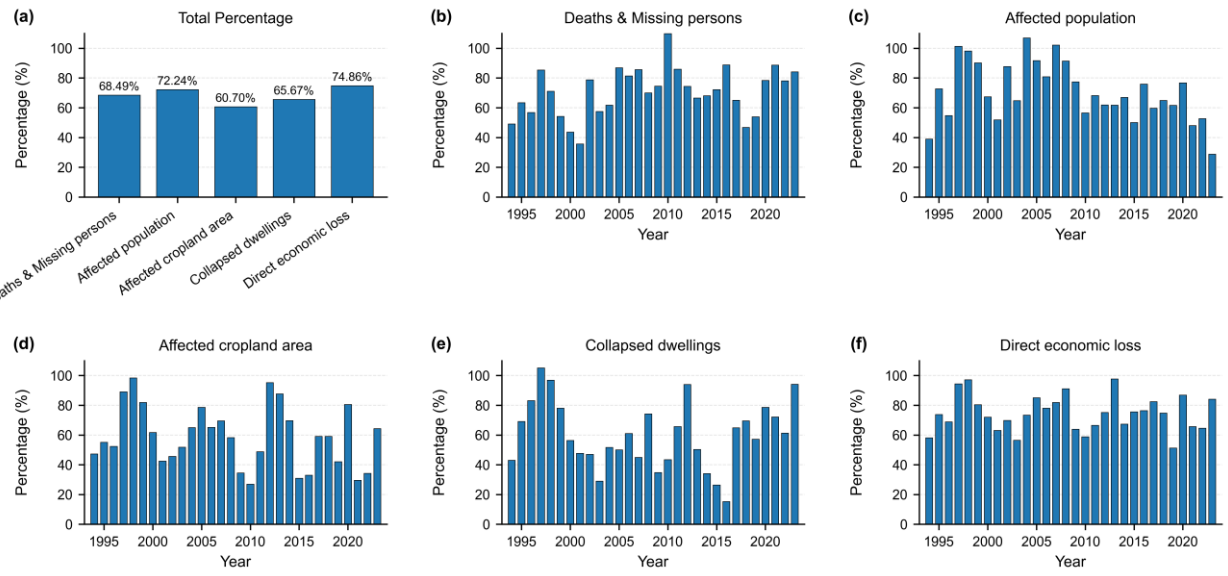


Figure 2. Percentages of official national totals represented in CFDID for five core flood impact indicators. Percentages are calculated using L1 records only. (a) Total percentages during 1994–2023. (b–f) Annual percentages for deaths and missing persons, affected population, affected cropland area, collapsed dwellings, and direct economic losses.

The annual percentages vary substantially among indicators and years (Fig. 2b–f). The annual percentage ranges from 35.70% to 109.86% for deaths and missing persons, from 28.83% to 107.00% for affected population, from 26.98% to 98.41% for affected cropland area, from 15.25% to 105.19% for collapsed dwellings, and from 51.22% to 97.70% for direct economic

losses. Some annual percentages exceed 100%, which may reflect differences in statistical approaches between event-scale records and official annual totals. Similar discrepancies are also observed in EM-DAT when event-level records are compared with official annual totals (Fig. 7)."

4.1 Overall comparison with international flood impact databases: Lines 324-331:

"The international databases each have their own strengths and can complement CFDID. For deaths and missing persons, EM-DAT accounts for a slightly higher percentage of the corresponding official total than CFDID. DFO provides spatial information derived from remote sensing and is valuable for locating high-impact flood events and characterizing affected or inundated areas. Wikimpacts provides a rich set of impact categories, although the percentages of the corresponding official totals represented in its China-related records remain relatively low at present. This may partly reflect the limited representation of Chinese-language flood-related news and reports in Wikipedia-based sources (Li et al., 2026).

Table 4. Comparison of event records and percentages of official national totals represented in CFDID and international flood impact databases during 1994–2023. CFDID is compared using L1 records, whereas Wikimpacts is compared using L2 national-level China records.

Indicator	CFDID (L1)	EM-DAT	DFO	Wikimpacts (L2)
Number of event records	1391	434	288	151
Deaths and missing persons	68.49%	72.15%	37.81%	17.12%
Affected population	72.24%	46.49%	/	38.21%
Affected cropland area	60.70%	/	/	/
Collapsed dwellings	65.67%	/	/	8.27%
Direct economic losses	74.86%	56.41%	/	38.14%

"

SC2:

Comment: *Minor figures corrections.*

I found many (which is why I prefer to report it as specific comments, rather than technical corrections) small differences in several figures in sections 3.1 and 3.4 (L202-245, e.g., I count

1418 records with affected population, not 1419; 88 L1 in 2005, not 94; ...). I suggest checking the counting method, making it scriptable if not yet the case, recomputing the statistics from the archived dataset, and carefully revising the numbers in the manuscript.

Response:

We sincerely thank you for your careful checking. We confirm that the number of L1 records in 2005 is 88, rather than 94 as stated in the original manuscript, and have corrected this value.

Regarding affected population, your count of 1,418 records was correct for the previously deposited dataset. The discrepancy arose because one affected-population value had been inadvertently omitted from that version. We have corrected this omission in the updated dataset on Zenodo. The revised database therefore contains 1,419 records with affected-population values, while the total number of event records remains 1,716. We apologize for the inconsistency between the manuscript and the previously deposited file.

To address these discrepancies and ensure consistency throughout the revised manuscript, we recomputed the descriptive statistics directly from the revised database. We systematically cross-checked the total record counts, the annual L1 record counts, the numbers of non-null values for each impact indicator, and the corresponding text, figures, and tables. We have revised the relevant statistics in Sects. 3.1 and 3.4 and updated the other affected results to ensure consistency with the revised dataset deposited on Zenodo.

3.1 Overview of CFDID: Lines 225-228:

"Among the 1716 records, 1205 records contain deaths and missing persons; 1419 records contain affected population; 1241 records contain affected cropland area; 87 records contain damaged cropland area; 395 records contain failed cropland area; 969 records contain collapsed dwellings; 397 records contain damaged dwellings; and 1444 records contain direct economic losses."

3.4 Temporal distribution: Lines 266-268:

"During 1994–2023, the largest number of records occurred in 2008, with 106 events, followed by 2009 (102 events), 2007 (94 events) and 2005 (88 events)."

SC3:

Comment: *State the filtering details and assumptions behind the comparison with international databases*

- Which EM-DAT and Wikimpacts disaster types were counted. The selection is not stated, and it materially changes the comparison: flood-only and flood-plus-tropical-cyclone give very different totals for China.

- Which version of each database was used, another important disclosure for reproducibility. Could also be part of section 5.

- Whether entry criteria were matched. EM-DAT admits an event only if it meets at least one threshold — ⩾10 deaths, ⩾100 affected, a declared state of emergency, or a call for international assistance. The author might consider a fair comparison, though it will not map the CFDID actual content. I would suggest at least disclosing the number of events and impact shares that would not be reported in EM-DAT due to the entry-criteria gate.

Response:

We sincerely thank you for this constructive comment. We agree that the filtering criteria, database versions, and assumptions underlying the comparison with international databases should be stated more explicitly. We have therefore added these details to the revised manuscript and clarified the limitations of the comparison.

First, we clarified the disaster types included in each database. For EM-DAT, we included all records with Disaster Type = "Flood", as well as records whose Disaster Subtype was classified as "Landslide (wet)", "Rockfall (wet)", "Mudslide", "Tropical cyclone", or "Glacial lake outburst

flood". For Wikimpacts, we used L2 records associated with L1 main events classified as "Flood", "Tropical Storm/Cyclone", or "Extratropical Storm/Cyclone", consistent with the flood-related event scope adopted for CFDID.

Second, the EM-DAT data used in this study were downloaded on 28 April 2026. We used Wikimpacts 1.0 and DFO Global Flood Records v0.9.0. We have added the corresponding version information, data endpoints, and access dates to the revised manuscript to improve reproducibility in the Section 5.

Third, we examined the influence of EM-DAT's numerical entry thresholds by considering CFDID L1 records reporting at least 10 deaths and missing persons or at least 100 affected persons. CFDID does not systematically record declarations of emergency or calls for international assistance, and missing impact values cannot be treated as zero. We therefore assessed the two numerical conditions without attempting to reproduce all EM-DAT admission decisions.

Among the 1,391 CFDID L1 records, 1,246 records meet at least one numerical threshold. Of these, 540 have identified EM-DAT counterparts and 706 do not. The reported impacts of the latter records account for 13.67%, 19.95%, and 16.34% of the corresponding official national totals for deaths and missing persons, affected population, and direct economic losses during 1994–2023, respectively. Thus, even among records meeting at least one numerical threshold, CFDID provides substantial additional impact information through records without identified EM-DAT counterparts. We have incorporated this analysis into Sect. 4.2, alongside the magnitude-stratified comparison in Fig. 7d–f. For the remaining 145 CFDID L1 records (10.42%), the available information does not demonstrate that either numerical threshold is met. Together, these records report 519 deaths and missing persons, 20 affected persons, and CNY 24.99 billion in direct economic losses at 2023 prices. Their reported mortality and economic losses represent 0.99% and 0.35% of the corresponding official national totals, respectively. Excluding these

records therefore has little effect on the aggregate comparison, but they cannot be classified as confirmed failures of EM-DAT's full entry criteria.

4.1 Overall comparison with international flood impact databases: Lines 311-315:

"For EM-DAT, we included all records classified as "Flood", together with flood-related records corresponding to Landslide (wet), Rockfall (wet), Mudslide, Tropical cyclone, and Glacial lake outburst flood. For Wikimpacts, we used its L2 records associated with L1 main events classified as "Flood", "Tropical Storm/Cyclone", or "Extratropical Storm/Cyclone". Database versions and access information are provided in Sect. 5."

4.2 Detailed comparison with EM-DAT: Lines 332-361:

"Among the international databases considered, EM-DAT contains the largest number of event records and accounts for the highest percentages of the corresponding official totals for deaths and missing persons, affected population, and direct economic losses. We further compare CFDID and EM-DAT with official annual totals for deaths and missing persons, affected population, and direct economic losses (Fig. 7a–c). Overall, the three series show broadly consistent interannual variability, indicating that both databases capture major fluctuations in flood impacts. CFDID and EM-DAT show relatively similar annual patterns for deaths and missing persons, whereas CFDID reports higher affected population and direct economic losses than EM-DAT in most years.

Because annual totals can be dominated by large events, we further examined correspondence between CFDID L1 records and EM-DAT records using the matching procedure in Appendix C. The EM-DAT record "2018-0369-CHN" was excluded because its location information was insufficient, leaving 433 EM-DAT records for comparison with 1,391 CFDID L1 records. Figure 7d–f compares record counts across different magnitude classes, distinguishing records with and without identified counterparts. Missing values are displayed separately as N/A and are not treated as zero. In total, 822 CFDID records (59.09%) and 120 EM-DAT records (27.71%) have no identified counterpart. The reported impacts of the unmatched CFDID records

account for 14.43%, 19.95%, and 16.57% of the official national totals for deaths and missing persons, affected population, and direct economic losses during 1994–2023, respectively.

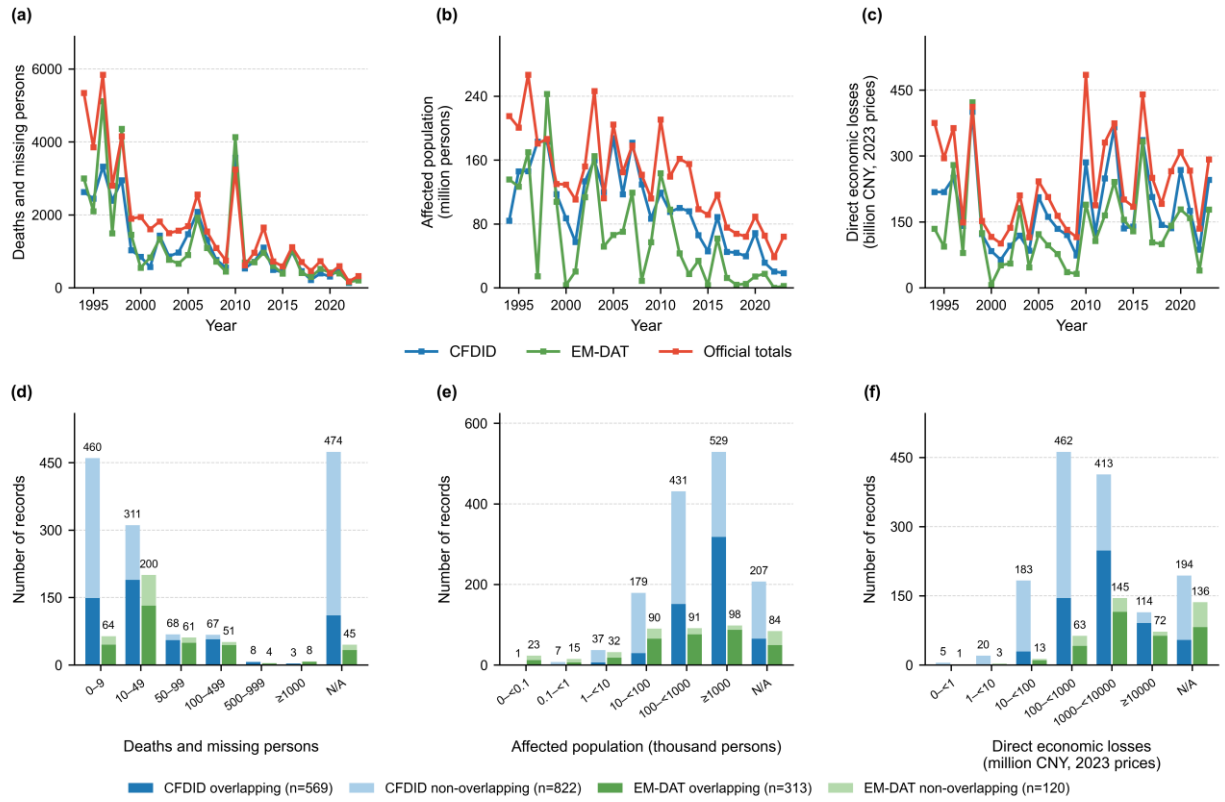


Figure 7. Comparison of flood impacts in CFDID, EM-DAT, and official national totals during 1994–2023. (a–c) Annual deaths and missing persons, affected population, and direct economic losses. (d–f) Record counts by magnitude, with darker and lighter segments indicating records with and without identified counterparts, respectively. Economic losses are expressed at 2023 prices.

Considering EM-DAT's two numerical entry criteria, 1,246 CFDID L1 records report at least 10 deaths and missing persons or at least 100 affected persons. Of these, 540 have identified EM-DAT counterparts and 706 do not. The reported impacts of the latter records account for 13.67%, 19.95%, and 16.34% of the official national totals for deaths and missing persons, affected population, and direct economic losses during 1994–2023, respectively. Conversely, 120

EM-DAT records have no identified CFDID counterpart. Their reported impacts represent 11.29%, 2.23%, and 5.09% of the corresponding official national totals.

These results indicate that CFDID and EM-DAT provide complementary information on flood impacts. Further work could harmonize the spatial and temporal scopes of matched record groups to assess the consistency and complementary detail of their reported impacts. The identifiers of matched and unmatched records are provided with the CFDID release on https://zenodo.org/records/22918521."

5 Data availability: Lines 409-411:

"EM-DAT is available at https://www.emdat.be (data export downloaded 28 April 2026). Wikimpacts 1.0 is available at https://bolin.su.se/data/li-2025-wikimpacts-1.0.final. DFO Global Flood Records v0.9.0 is available at https://doi.org/10.5281/zenodo.19288171."

Appendix C: Rules of matching CFDID and EM-DAT records: Lines 491-507:

"Record correspondences were assessed between 1,391 CFDID L1 records and 433 EM-DAT records. EM-DAT record 2018-0369-CHN was excluded because its location information was insufficient for matching. Candidate correspondences were identified using reported provinces, overlapping time intervals, and disaster categories. Typhoon records were examined separately from other flood-related records because their names or identification numbers generally allow more precise cross-checking.

For candidate screening, missing start and end days were represented by the first and last days of their respective months, provided that the corresponding year and month were known. All end-date fields were missing in 337 CFDID L1 records, representing 24.2% of the CFDID comparison sample, whereas no EM-DAT records in the comparison had all end-date fields missing. For these CFDID records, the candidate interval extended from the reported start date to the last day of the start month. When the start day was also missing, the interval began on the first

day of that month. These temporal bounds accommodate date uncertainty and were not interpreted as exact event dates.

Broad time windows and shared provinces were used only to identify candidate correspondences. All final correspondences were determined through record-by-record manual checking. One-to-many and many-to-many correspondences were allowed because the databases aggregate events differently. A record was classified as overlapping if at least one counterpart was identified. The identifiers of matched and unmatched records are provided in the CFDID repository on Zenodo https://doi.org/10.5281/zenodo.22918521."

SC4:

Comment: *Event count indicators are misleading.*

Event counts are reported in several places (e.g., Figure 4, Table 3), however, when it comes to social sensing of floods (i.e., sourced from societal sources with no control over what defines an event), there are no spatiotemporal criteria to frame the definition and compare event counts consistently. This fact is acknowledged by the authors for temporal trends at L247-249, and CFDID's own L1/L2/L3 hierarchy is itself an aggregation choice.

My point is that the same caution should be acknowledged when comparing different databases (Table 3). The number of events is more an indicator of the level of detail in a database than of coverage. In this case, however, it is possible to define aggregation-invariant indicators, e.g., the percentage of events that do not overlap with EM-DAT events, or to compare flood events based on overlapping days.

Response:

We sincerely thank you for this constructive comment. We agree that record counts are influenced by event delineation and aggregation and cannot alone establish database completeness. We have clarified this point in Sect. 4.1 and extended Sect. 4.2 to examine record correspondences and the reported impacts associated with records without identified counterparts.

We compared 1,391 CFDID L1 records with 433 EM-DAT records. The EM-DAT record 2018-0369-CHN is excluded, because its location information was insufficient for matching. Correspondences were assessed using reported provinces, time intervals, and disaster categories, followed by record-by-record manual checking. One-to-many and many-to-many correspondences were allowed to accommodate differences in event aggregation. The procedure is described in Appendix C.

In total, 822 CFDID records (59.09%) and 120 EM-DAT records (27.71%) have no identified counterpart. The reported impacts of the unmatched CFDID records account for 14.43%, 19.95%, and 16.57% of the official national totals for deaths and missing persons, affected population, and direct economic losses during 1994–2023, respectively. Conversely, the unmatched EM-DAT records account for 11.29%, 2.23%, and 5.09% of the corresponding official totals. These results quantify complementary impact information in both databases rather than relying solely on differences in their total record counts.

We acknowledge that records without identified counterparts are not necessarily disasters omitted from the other database, because the results remain sensitive to event aggregation and the information available for matching. The present analysis identifies record correspondences; further work could harmonize the spatial and temporal scopes of matched record groups to assess the consistency and complementary detail of their reported impacts. The identifiers of matched and unmatched records are provided with the revised CFDID release on Zenodo.

4.2 Detailed comparison with EM-DAT: Lines 332-361:

“Among the international databases considered, EM-DAT contains the largest number of event records and accounts for the highest percentages of the corresponding official totals for deaths and missing persons, affected population, and direct economic losses. We further compare CFDID and EM-DAT with official annual totals for deaths and missing persons, affected population, and direct economic losses (Fig. 7a–c). Overall, the three series show broadly consistent interannual variability, indicating that both databases capture major fluctuations in flood impacts. CFDID and

EM-DAT show relatively similar annual patterns for deaths and missing persons, whereas CFDID reports higher affected population and direct economic losses than EM-DAT in most years.

Because annual totals can be dominated by large events, we further examined correspondence between CFDID L1 records and EM-DAT records using the matching procedure in Appendix C. The EM-DAT record "2018-0369-CHN" was excluded because its location information was insufficient, leaving 433 EM-DAT records for comparison with 1,391 CFDID L1 records. Figure 7d–f compares record counts across different magnitude classes, distinguishing records with and without identified counterparts. Missing values are displayed separately as N/A and are not treated as zero. In total, 822 CFDID records (59.09%) and 120 EM-DAT records (27.71%) have no identified counterpart. The reported impacts of the unmatched CFDID records account for 14.43%, 19.95%, and 16.57% of the official national totals for deaths and missing persons, affected population, and direct economic losses during 1994–2023, respectively.

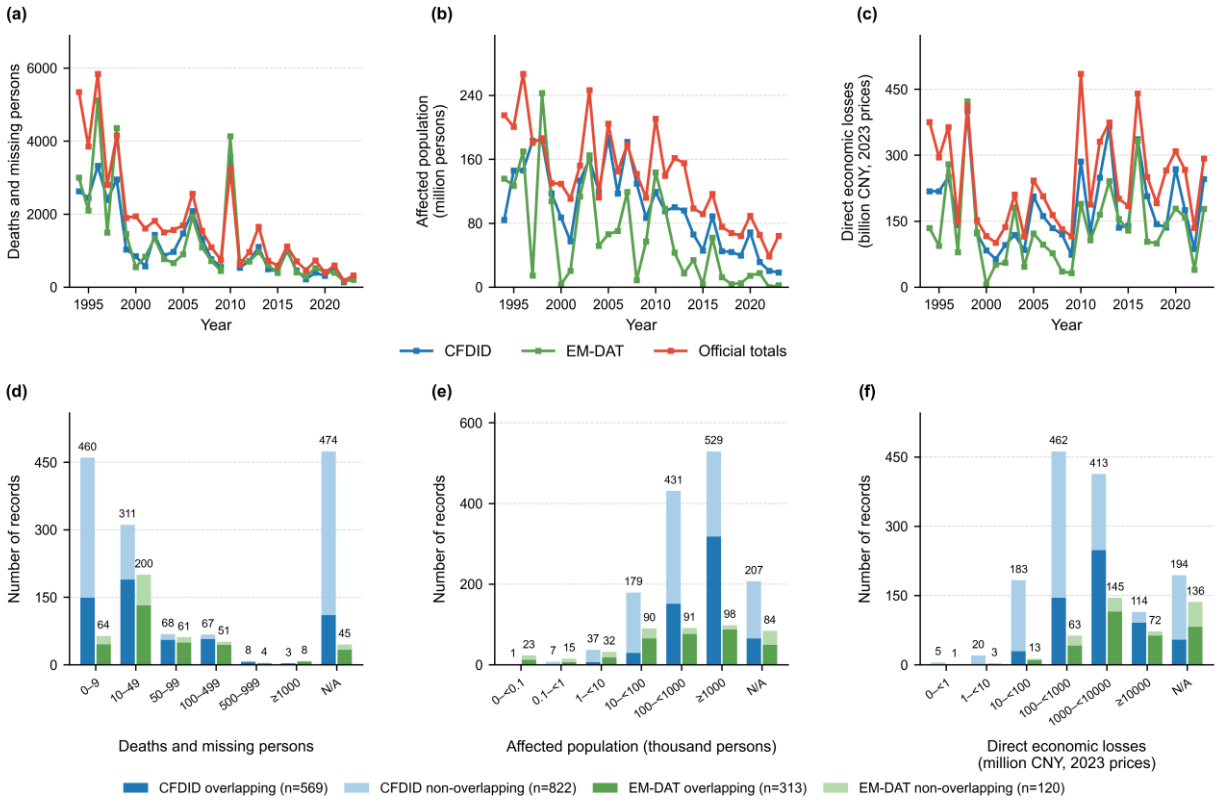


Figure 7. Comparison of flood impacts in CFDID, EM-DAT, and official national totals during 1994–2023. (a–c) Annual deaths and missing persons, affected population, and direct economic losses. (d–f) Record counts by magnitude, with darker and lighter segments indicating records with and without identified counterparts, respectively. Economic losses are expressed at 2023 prices.

Considering EM-DAT's two numerical entry criteria, 1,246 CFDID L1 records report at least 10 deaths and missing persons or at least 100 affected persons. Of these, 540 have identified EM-DAT counterparts and 706 do not. The reported impacts of the latter records account for 13.67%, 19.95%, and 16.34% of the official national totals for deaths and missing persons, affected population, and direct economic losses during 1994–2023, respectively. Conversely, 120 EM-DAT records have no identified CFDID counterpart. Their reported impacts represent 11.29%, 2.23%, and 5.09% of the corresponding official national totals.

These results indicate that CFDID and EM-DAT provide complementary information on flood impacts. Further work could harmonize the spatial and temporal scopes of matched record groups to assess the consistency and complementary detail of their reported impacts. The identifiers of matched and unmatched records are provided with the CFDID release on https://doi.org/10.5281/zenodo.22918521."

Appendix C: Rules of matching CFDID and EM-DAT records: Lines 491-507:

"Record correspondences were assessed between 1,391 CFDID L1 records and 433 EM-DAT records. EM-DAT record 2018-0369-CHN was excluded because its location information was insufficient for matching. Candidate correspondences were identified using reported provinces, overlapping time intervals, and disaster categories. Typhoon records were examined separately from other flood-related records because their names or identification numbers generally allow more precise cross-checking.

For candidate screening, missing start and end days were represented by the first and last days of their respective months, provided that the corresponding year and month were known. All end-date fields were missing in 337 CFDID L1 records, representing 24.2% of the CFDID comparison sample, whereas no EM-DAT records in the comparison had all end-date fields missing. For these CFDID records, the candidate interval extended from the reported start date to the last day of the start month. When the start day was also missing, the interval began on the first day of that month. These temporal bounds accommodate date uncertainty and were not interpreted as exact event dates.

Broad time windows and shared provinces were used only to identify candidate correspondences. All final correspondences were determined through record-by-record manual checking. One-to-many and many-to-many correspondences were allowed because the databases aggregate events differently. A record was classified as overlapping if at least one counterpart was identified. The identifiers of matched and unmatched records are provided in the CFDID repository on Zenodo https://doi.org/10.5281/zenodo.22918521. ”

SC5:

Comment: *The text concludes that CFDID, EM-DAT and the official totals “show broadly consistent interannual variability, indicating that both databases capture major fluctuations in flood impacts” (L307). However, disaster impact distributions are highly concentrated (Pareto-distributed). For instance, the CFDID top 10% deadliest events account for 66% of mortality in the database. Annual totals are therefore mostly driven by a handful of large events that are likely present in the other databases as well. Agreement (or correlation between Fig.7’s series) at that level is close to guaranteed, and says little about the rest of the record. This matters because the paper's own argument for CFDID's added value lies exactly where the databases should differ: at l. 34–35, EM-DAT is said to omit small-scale and localized events. The paper would be improved, in my opinion, if the authors provided statistics that explicitly support the implicitly claimed added value of CFDID (in line with SC4). Comparing the databases in the body of the distribution*

rather than in the total (for example, after excluding the largest events) or comparing counts and impacts below a magnitude threshold could do this.

Response:

We sincerely thank you for this constructive comment. We agree that annual totals can be dominated by a small number of large events and that similarity in annual totals alone does not adequately characterize differences across the event-level impact distribution.

Following your suggestion, we have extended the comparison beyond the annual series in Fig. 7a–c. Figure 7d–f now presents record counts across different magnitude classes for deaths and missing persons, affected population, and direct economic losses. Within each class, the counts distinguish records with and without identified counterparts in the other database. Missing values are displayed separately as N/A and are not treated as zero. These panels allow record counts and overlap status in lower and intermediate magnitude classes to be examined separately from the largest-impact class.

The revised Sect. 4.2 combines these magnitude-stratified comparisons with the manually checked record correspondences and quantifies the reported impacts associated with records without identified counterparts in both databases. It also examines this complementary information among CFDID records meeting at least one of EM-DAT's two numerical entry thresholds. We therefore discuss complementarity on the basis of record-level comparisons and their reported impacts, rather than interpreting agreement in annual totals as evidence of database completeness.

4.2 Detailed comparison with EM-DAT: Lines 332-361:

"Among the international databases considered, EM-DAT contains the largest number of event records and accounts for the highest percentages of the corresponding official totals for deaths and missing persons, affected population, and direct economic losses. We further compare CFDID and EM-DAT with official annual totals for deaths and missing persons, affected population, and

direct economic losses (Fig. 7a–c). Overall, the three series show broadly consistent interannual variability, indicating that both databases capture major fluctuations in flood impacts. CFDID and EM-DAT show relatively similar annual patterns for deaths and missing persons, whereas CFDID reports higher affected population and direct economic losses than EM-DAT in most years.

Because annual totals can be dominated by large events, we further examined correspondence between CFDID L1 records and EM-DAT records using the matching procedure in Appendix C. The EM-DAT record "2018-0369-CHN" was excluded because its location information was insufficient, leaving 433 EM-DAT records for comparison with 1,391 CFDID L1 records. Figure 7d–f compares record counts across different magnitude classes, distinguishing records with and without identified counterparts. Missing values are displayed separately as N/A and are not treated as zero. In total, 822 CFDID records (59.09%) and 120 EM-DAT records (27.71%) have no identified counterpart. The reported impacts of the unmatched CFDID records account for 14.43%, 19.95%, and 16.57% of the official national totals for deaths and missing persons, affected population, and direct economic losses during 1994–2023, respectively.

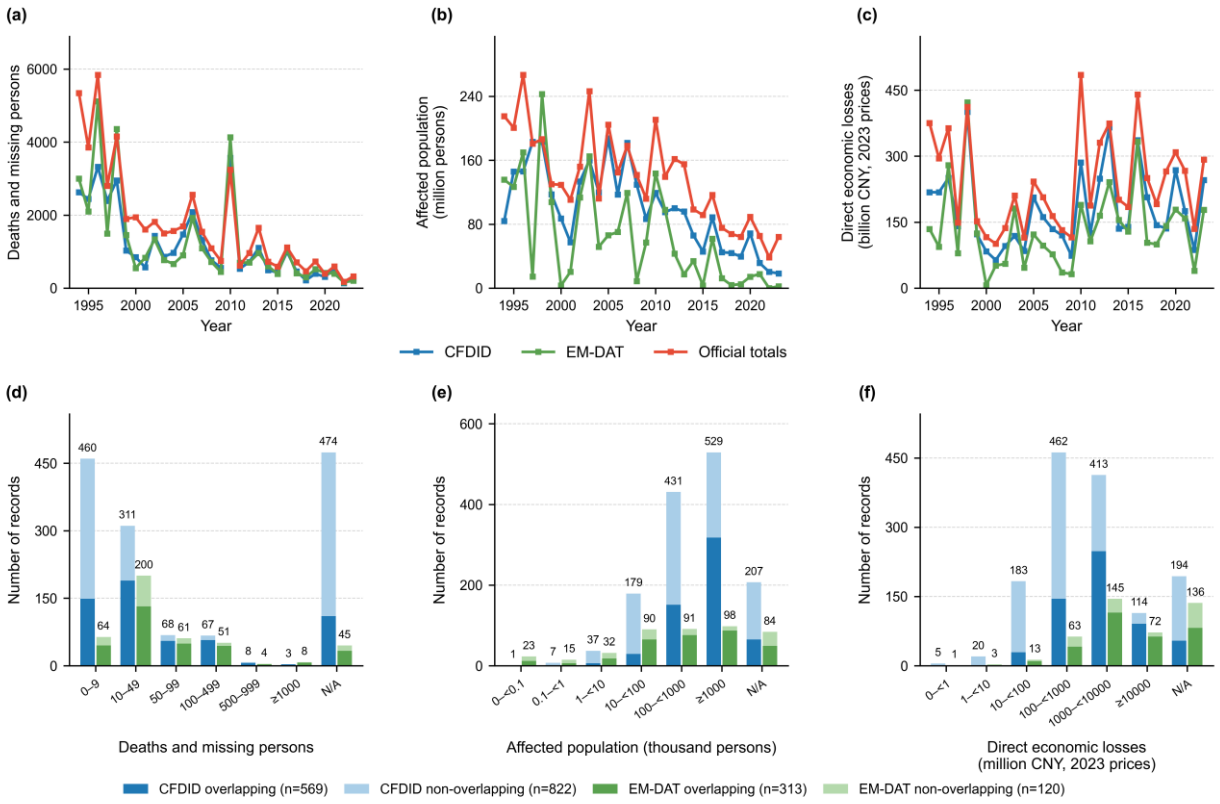


Figure 7. Comparison of flood impacts in CFDID, EM-DAT, and official national totals during 1994–2023. (a–c) Annual deaths and missing persons, affected population, and direct economic losses. (d–f) Record counts by magnitude, with darker and lighter segments indicating records with and without identified counterparts, respectively. Economic losses are expressed at 2023 prices.

Considering EM-DAT's two numerical entry criteria, 1,246 CFDID L1 records report at least 10 deaths and missing persons or at least 100 affected persons. Of these, 540 have identified EM-DAT counterparts and 706 do not. The reported impacts of the latter records account for 13.67%, 19.95%, and 16.34% of the official national totals for deaths and missing persons, affected population, and direct economic losses during 1994–2023, respectively. Conversely, 120 EM-DAT records have no identified CFDID counterpart. Their reported impacts represent 11.29%, 2.23%, and 5.09% of the corresponding official national totals.

These results indicate that CFDID and EM-DAT provide complementary information on flood impacts. Further work could harmonize the spatial and temporal scopes of matched record groups to assess the consistency and complementary detail of their reported impacts. The identifiers of matched and unmatched records are provided with the CFDID release on https://doi.org/10.5281/zenodo.22918521."

Appendix C: Rules of matching CFDID and EM-DAT records: Lines 491-507:

"Record correspondences were assessed between 1,391 CFDID L1 records and 433 EM-DAT records. EM-DAT record 2018-0369-CHN was excluded because its location information was insufficient for matching. Candidate correspondences were identified using reported provinces, overlapping time intervals, and disaster categories. Typhoon records were examined separately from other flood-related records because their names or identification numbers generally allow more precise cross-checking.

For candidate screening, missing start and end days were represented by the first and last days of their respective months, provided that the corresponding year and month were known. All end-date fields were missing in 337 CFDID L1 records, representing 24.2% of the CFDID comparison sample, whereas no EM-DAT records in the comparison had all end-date fields missing. For these CFDID records, the candidate interval extended from the reported start date to the last day of the start month. When the start day was also missing, the interval began on the first day of that month. These temporal bounds accommodate date uncertainty and were not interpreted as exact event dates.

Broad time windows and shared provinces were used only to identify candidate correspondences. All final correspondences were determined through record-by-record manual checking. One-to-many and many-to-many correspondences were allowed because the databases aggregate events differently. A record was classified as overlapping if at least one counterpart was identified. The identifiers of matched and unmatched records are provided in the CFDID repository on Zenodo https://doi.org/10.5281/zenodo.22918521."

SC6:

Comment: *References and potential data sources exhaustiveness.*

Three dataset references are mentioned in L336. The differences between CFDID and these datasets should be discussed or introduced more clearly. This is a gap in the paper context. The China Flood and Drought Bulletin is not mentioned anywhere in the manuscript and does not appear in Table 1. Fu et al. (2025), one of these three references, benchmarks its own dataset against this Bulletin throughout and describes it as offering "authoritative data on flood disasters, focusing on economic losses, casualties, and agricultural damages at the provincial level". I don't have sufficient knowledge or language skills to evaluate whether this source should be used, and I leave that responsibility to the authors. Likewise, the paper may lack references and context on LLM-based impact extraction and refers to Li et al. (2026) as the sole example. A few more

references and a bit more context would help better anchor the paper's methodology in existing work.

Response:

We sincerely thank you for this constructive comment. We agree that the original manuscript did not provide sufficient context on existing China-focused flood datasets or on previous applications of large language models for disaster information extraction.

First, regarding the China Flood and Drought Bulletin, the flood-disaster statistics relevant to this study are the same as those reported in the China Water Resources Yearbook, and both are compiled within the Chinese water resources reporting system. Because the China Water Resources Yearbook is already one of the core sources used to construct CFDID, we did not treat the Bulletin as an additional independent data source. We have clarified this point in Sect. 2.2.

Second, we have expanded the discussion of the datasets presented in Fu et al. (2025), Shen et al. (2025), and Qiu et al. (2025), and clarified their differences from and complementarity with CFDID. Fu et al. (2025) constructed a county-level urban flood-event dataset for China from news reports, while Shen et al. (2025) identified urban flood events from Sina Weibo. These datasets provide valuable information on local flood occurrence but do not systematically include the range of impact indicators contained in CFDID. Qiu et al. (2025), in contrast, used Sentinel-1 remote sensing to map flooded croplands at high spatial resolution, providing spatially explicit information that can complement the administrative-unit-based records in CFDID. We have added this discussion to Sect. 4.4 and highlighted these datasets as potential sources for future expansion and cross-validation of CFDID.

Third, we have added additional references on the use of LLMs for disaster information extraction. Previous studies have demonstrated the potential of LLMs to extract disaster events and impact information from digital text sources. Our study does not present the LLM itself as a methodological innovation; rather, we develop an integrated workflow combining OCR, rule-

based text segmentation, LLM-based information extraction, and manual verification for historical printed official documents. The workflow is intended to improve the efficiency and reproducibility of structuring such historical records, while the primary contribution of this study remains the manually verified CFDID database itself. We have clarified this point and added the relevant references in Sect. 2.4.3.

2.2 Data sources: Lines 111-114:

"Another official source of flood-disaster statistics is the China Flood and Drought Bulletin, published by the Chinese water resources authorities. The flood-disaster statistics relevant to this study are the same as those reported in the China Water Resources Yearbook. Because the China Water Resources Yearbook has already been systematically incorporated into CFDID, we did not treat the Bulletin as an additional independent data source."

2.4.3 Data extraction: Lines 158-161:

"Recent studies have demonstrated the potential of LLMs for extracting disaster events and impact information from unstructured or semi-structured text sources (Hasbini et al., 2026; Li et al., 2026). Building on these advances, we applied an LLM-based extraction procedure to historical Chinese official documents after OCR conversion and rule-based text segmentation."

4.4 Limitations and future improvements: Lines 383-390:

"Additional datasets could provide complementary information for future extensions of CFDID. Fu et al. (2025) constructed a county-level urban flood-event dataset for 2000–2022 from news reports, while Shen et al. (2025) identified urban flood events for 2012–2023 using Sina Weibo. These datasets provide finer-scale information on flood occurrence, although they do not systematically report the range of disaster impact indicators. Remote-sensing products can provide complementary spatial information. For example, Qiu et al. (2025) mapped flooded croplands in China during 2017–2021 using Sentinel-1 observations at a much finer spatial resolution than administrative-unit-based disaster records. Physically derived hazard datasets, such as the Global

Drought and Flood Catalogue (He et al., 2020), can also provide independent information on event timing and spatial extent."

SC7:

Comment: *Complete the data and code repository*

A few things are missing to better align the data repository with FAIR principles, ESSD requirements, and make the reviewer's work easier: (i) The official annual national totals for the five core indicators are the denominator of every "coverage" figure in the paper but are not deposited, which impedes paper reproducibility. (ii) The repository has no README and no data dictionary, so Table 2's field definitions, units, and null semantics exist only in the paper. This is a restriction for machine readability. Besides, (iii) no machine-readable schema or validation script accompanies the data, so conformity with Table 2 and with the retention rule stated at L178–179 cannot be confirmed without writing one, which is especially important for a database intended to be updated annually. (iv) No requirements.txt or stated Python version. A short dependency list would make the code runnable. (v) Lastly, the ESSD paper DOI is not registered as a related identifier on the Zenodo record, with creator ORCIDs and affiliations also absent, with no corresponding author or contact point.

Response:

We sincerely thank you for this constructive comment. We agree that making the repository more self-contained and machine-readable is important for improving adherence to FAIR principles, reproducibility, and maintainability of CFDID. Following your suggestions, we have substantially expanded the Zenodo repository.

(i) We have deposited the official annual national totals for the five core flood-impact indicators during 1994–2023. These are the benchmark data used in Sect. 2.5, Fig. 2, and the

comparisons with international disaster databases, allowing users to reproduce the percentages relative to the official national totals reported in the manuscript.

(ii) We have added a README that describes the repository structure, individual files, database structure, event hierarchy, field definitions, units, missing-value semantics, data sources, construction workflow, and version-update policy. We have also provided a machine-readable data dictionary corresponding to the field definitions in Table 2.

(iii) We have added a database validation script that can be applied directly to the released CFDID file. The script checks the database structure and key consistency rules, including required fields, Event ID uniqueness and hierarchical relationships, valid dates and disaster categories, non-negative impact values, and the requirement that each retained record contain at least one value among the five core impact indicators. We also re-examined the database-construction workflow and confirmed that this retention rule is implemented during the data-extraction stage in 3_Data_Extraction.py. The original manuscript incorrectly described it as part of post-processing. We have therefore revised Sects. 2.4.3 and 2.4.5 so that the manuscript description is consistent with the released code.

(iv) We have added software-environment information to the repository. The workflow was run using Python 3.13.5, and a requirements.txt file is now provided to document the required Python dependencies.

(v) We have updated the Zenodo metadata to include the authors' ORCIDs and affiliations, as well as the corresponding-author contact information. We have also registered the available ESSD discussion/preprint DOI as a related identifier. The final article DOI will be added to the Zenodo record after formal publication.

2.4.3 Data extraction: Lines 164-166:

"Only event records with a positive value for at least one of the following five core impact indicators were retained, including deaths and missing persons, affected population, affected cropland area, collapsed dwellings, and direct economic losses."

5 Data availability: Lines 408:

"The repository also contains the official annual national totals used as benchmarks."

Other revisions can be found in the updated Zenodo (doi.org/10.5281/zenodo.22918521).

TC1:

Comment: *Table 2: "semi-required" is not standard terminology. Consider "conditional".*

Response:

We sincerely thank you for this suggestion. We agree that "semi-required" is not standard terminology. We have replaced all occurrences of "semi-required" with "conditional" in the manuscript and Table 2, and revised the corresponding description of field requirement levels in Sect. 2.3 accordingly.

2.3 Database fields: Lines 122-130:

"Table 2 lists the field names, descriptions, data types, units, and requirement levels in the released database. The "Required" column indicates whether a field is mandatory, conditional, or optional. A "Yes" field must be recorded for each event record. The "conditional" typhoon field means that the typhoon ID or name should be recorded when the disaster category is typhoon. The "conditional" impact fields do not mean that every event must contain all impact indicators; rather, a record must contain at least one of the five core impact indicators: deaths and missing persons, affected population, affected cropland area, collapsed dwellings, or direct economic losses. Fields marked as "No" are optional and are recorded only when the original source provides the corresponding information. Missing values are recorded as null, indicating that the original source did not report the information; they should not be interpreted as zero."

Table 2. All fields of CFDID.

No.	Field Name	Description	Data type	Unit	Required
1	Event level	Hierarchical record level. L1 represents a parent-event record, L2 represents a sub-event record nested within an L1 record, and L3 represents a more detailed record nested within an L2 record.	Categorical	/	Yes
2	Event ID	Unique identifier of each event record. L1 records use "YYYY-NNN"; L2 records use "YYYY-NNN-SS"; and L3 records use "YYYY-NNN-SS-GG", where "NNN", "SS", and "GG" indicate the event, sub-event, and sub-sub-event sequence numbers, respectively.	String	/	Yes
3	Start year	Year when flood impacts began.	Integer	/	Yes
4	Start month	Month when flood impacts began.	Integer	/	Yes
5	Start day	Day when flood impacts began.	Integer	/	No
6	End year	Year when direct impacts ended.	Integer	/	No
7	End month	Month when direct impacts ended.	Integer	/	No
8	End day	Day when direct impacts ended.	Integer	/	No
9	Province	31 provinces in the Chinese mainland; Hong Kong, Macao, and Taiwan are not included.	List	/	Yes
10	Specific location	Original place names such as cities, counties, townships, villages, or river basins.	String	/	No
11	Disaster category	Rainstorm flood, typhoon, snowmelt flood, ice-jam flood, or dam-break flood.	Categorical	/	Yes
12	Typhoon information	Typhoon number or name.	String	/	Conditional
13	Deaths and missing persons	Number of people killed or missing as a direct result of the flood disaster.	Non-negative integer	persons	Conditional
14	Affected population	Number of people directly affected by the disaster.	Non-negative decimal	thousand persons	Conditional
15	Affected cropland area	Sown area of crops with yield reduction of more than 10%.	Non-negative decimal	thousand ha	Conditional
16	Damaged cropland area	Sown area of crops with yield reduction of more than 30%.	Non-negative decimal	thousand ha	No
17	Failed cropland area	Sown area of crops with yield reduction of more than 80%.	Non-negative decimal	thousand ha	No
18	Collapsed dwellings	Number of dwelling rooms that collapsed or were severely damaged and required reconstruction.	Non-negative decimal	thousand rooms	Conditional
19	Damaged dwellings	Number of dwelling rooms structurally damaged but repairable.	Non-negative decimal	thousand rooms	No
20	Direct economic losses	Direct economic losses reported at current prices.	Non-negative decimal	million CNY	Conditional
21	Direct economic losses (2023 constant prices)	Direct economic losses adjusted to 2023 prices using CPI.	Non-negative decimal	million CNY	Conditional
22	Data source	Source document(s) from which the information was obtained.	List	/	Yes

TC2:

Comment: *Figure 2: There is a rounding mismatch between the figures on the top of the bars in panel (a) and the figures mentioned in the text at L215.*

Response:

We sincerely thank you for the careful checking. We recalculated and cross-checked the values reported in Fig. 2a and the corresponding text and standardized all values to two decimal places.

3.2 Percentages relative to official annual totals: Lines 235-243:

"The L1 records in CFDID represent substantial percentages of the official national totals for the five core impact indicators. Over 1994–2023, the corresponding total percentages are 68.49% for deaths and missing persons, 72.24% for affected population, 60.70% for affected cropland area, 65.67% for collapsed dwellings, and 74.86% for direct economic losses (Fig. 2a). Comparisons with international disaster impact databases are provided in Sect. 4.1.

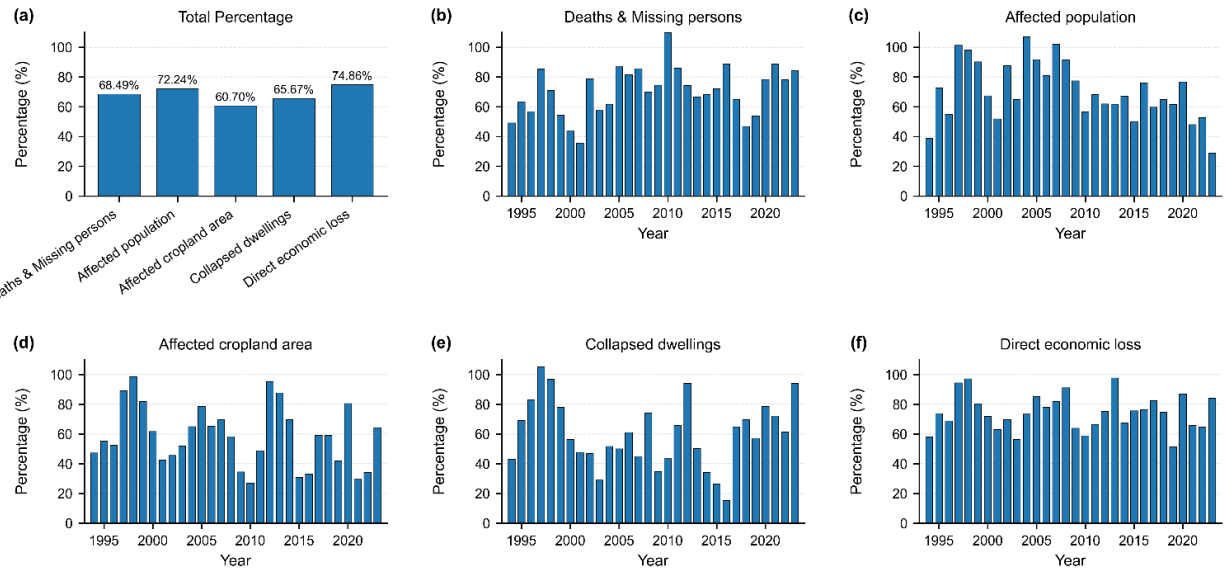


Figure 2. Percentages of official national totals represented in CFDID for five core flood impact indicators. Percentages are calculated using L1 records only. (a) Total percentages during 1994–2023. (b–f) Annual percentages for deaths and missing persons, affected population, affected cropland area, collapsed dwellings, and direct economic losses."

TC3:

Comment: *L340-343 present the conversion of known spatial information into spatial layers as future work, but Fig. 6 already maps L1 record counts onto all 31 provinces, so the provincial join has evidently been made. Consider narrowing the sentence to the levels that genuinely remain (prefectures, counties, river basins, and precise coordinates) and stating how the provincial join was performed.*

Response:

We sincerely thank you for this constructive comment. We agree that our previous wording was imprecise because province-level spatial information has already been used to produce Fig. 6. We have revised this paragraph to clarify how the provincial-level map was generated and to narrow the proposed future spatial development to finer spatial units and more precise geographic information.

Specifically, for Fig. 6, each L1 event was assigned to all provinces listed in its "Province" field, and the corresponding province-level event counts were then mapped. We have also clarified that future work will focus on disaggregating impacts of multi-province events and converting finer spatial information, such as prefectures, counties, and river basins, into spatial layers.

4.4 Limitations and future improvements: Lines 391-396:

"Second, province-level spatial information has already been incorporated into Fig. 6. Each L1 event was assigned to all provinces listed in its "Province" field, and the resulting province-level event counts were mapped accordingly. However, for events affecting multiple provinces, the reported impacts cannot yet be fully disaggregated among individual provinces. Future work will therefore focus on finer-scale spatial disaggregation and on converting information recorded in the "Specific location" field, including prefectures, counties, and river basins, into spatial layers

(Rosvold and Buhaug, 2021; Teber et al., 2026). This would facilitate more detailed spatial analyses and integration with hydrological, socioeconomic, and remote-sensing datasets."

TC4:

Comment: *Record 2007-083: reports only damaged dwellings (i.e., 8), which is not among the five core indicators of L178–179. Either the rule needs rewording, or the record should be removed.*

Response:

We sincerely thank you for the careful checking. We re-examined the original source for Event 2007-083 and found that the event also reported a direct economic loss of CNY 1 million, which had been inadvertently omitted during database compilation. We have added this value to the database. The record therefore now contains one of the five core impact indicators defined in Sect. 2.4.3 and satisfies the stated retention rule.

Because this correction affects the aggregated direct economic loss statistics, we have also recalculated and updated all corresponding direct economic loss results in the database, manuscript, figures, and tables.

TC5:

Comment: *Record 2006-068: carries an end date (15 September) earlier than its start date (19 September). One of the two is presumably a typo.*

Response:

We sincerely thank you for the careful checking. We rechecked the original source for Event 2006-068 and confirmed that the correct end date is 19 September, the same as the start

date. The previously recorded end date of 15 September was a transcription error. We have corrected the end date to 19 September in the revised database.

TC6:

Comment: *Consider adding one more decimal place to the two Direct economic losses columns; for small losses, the adjustment is not apparent due to rounding (e.g., 2011-001 or 2011-028, can be checked with an equal test).*

Response:

We sincerely thank you for this helpful suggestion. We have increased the numerical precision of both direct economic loss fields—the current-price losses and the losses adjusted to 2023 constant prices—to three decimal places. This allows the CPI adjustment to remain visible for events with relatively small losses, including Events 2011-001 and 2011-028 noted by the reviewer.