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

We sincerely thank all the 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. Data sources and event hierarchy. We have clarified the reporting scopes of the three official sources, including the sources of moderate and minor flood events. We have added an example illustrating the L1–L3 hierarchy and explained why CFDID L1 records and Wikimpacts L2 records are used in the comparison. Appendix A has also been expanded with an original event-description excerpt, its English translation, and the corresponding CFDID records.

2. Database construction workflow. We have expanded the description of within-source checking and cross-source fusion, specifying the candidate-screening conditions, LLM-assisted assessment of record relationships, and subsequent manual review. Figure 1 has been revised to show the main processing operations and intermediate outputs more explicitly.

3. Database comparisons and manuscript organization. We have clarified the scope of Table 4, explained the relevance of HANZE and Groundsource, and expanded the annual and record-level comparisons with EM-DAT. The revised analysis reports interannual correlations, bidirectional record correspondence, and the reported impacts of records without identified counterparts to clarify complementary reporting. The comparison results have been moved to the Results section as Sects. 3.6 and 3.7, while potential applications and limitations remain in the Discussion as Sects. 4.1 and 4.2.

4. Language revisions. We have refined the abstract, introduced all eight impact indicators together, and clarified the wording concerning parent-event records and percentages relative to official national totals. We have also corrected the Table 1 heading to "Number of documents".

We hope that these revisions address the reviewers' concerns and improve the clarity and usefulness of the manuscript. We sincerely thank the editors and reviewers again for their time and valuable suggestions. Detailed, point-by-point responses and the corresponding manuscript revisions are provided below.

Reviewer 1

R1C0:

Comment: *I appreciate the authors' thorough efforts in addressing my comments. The manuscript has been significantly improved, and I am pleased to recommend it for acceptance.*

Response:

We sincerely thank you for your positive assessment and recommendation for acceptance. We also appreciate your careful review and constructive suggestions throughout the review process, which have greatly helped us improve the manuscript.

Reviewer 3

R3C0:

General comments: *This study compiles the first China-based flood disaster event records. I agree with the previous reviewers that the work is valuable and deserves quick publication. These data will become key resources to support disaster analysis, model validation, and to guide policymakers and stakeholders. It is also a good supplement to current global-scale datasets (e.g., EM-DAT and DFO). After reviewing the comments, replies, and revisions in detail, I believe most of the reviewers' concerns are resolved, and the paper can be accepted. A few revisions remain before final acceptance.*

Response:

We sincerely thank you for your favorable assessment of the manuscript and for recognizing the value of CFDID. Your constructive suggestions have helped us further improve the manuscript. The main revisions are summarized below:

1. Language revisions. We have refined the abstract, introduced all eight impact indicators together, and corrected the Table 1 heading to "Number of documents" (R3C1 and R3C3).

2. Event hierarchy and illustrative examples. We have added a September 2011 example to illustrate the nested relationships among L1, L2, and L3 records, and clarified why the comparison uses CFDID L1 records and Wikimpacts L2 records. We have also expanded Appendix A with an excerpt from the 2023 Yearbook of Meteorological Disasters in China describing two events in Jilin in 2022, together with an English translation and the corresponding CFDID records (R3C2 and R3C4).

3. Data fusion procedures. We have clarified the distinction between within-source checking and cross-source fusion, specified the candidate-screening conditions, and explained the subsequent LLM-assisted relationship assessment and manual verification. These revisions

provide a more explicit description of how record relationships are assessed during database construction (R3C5).

4. EM-DAT comparison and figure presentation. We have corrected the order of the first citation to Fig. 7 and expanded the annual and record-level comparisons with EM-DAT. The revised analysis reports annual correlation coefficients, bidirectional record correspondence, and the reported impacts of records without identified counterparts relative to official national totals (R3C6 and R3C7).

We sincerely thank you again for your careful review and constructive suggestions. Detailed, point-by-point responses and the corresponding manuscript revisions are provided below.

R3C1:

Comment: *1. The language in the abstract can be improved.*

e.g., China is one of the countries …. Flood disaster worldwide. "worldwide" can be deleted.

e.g., L1 is not "top-level", as it should be explained as the parent-event records.

e.g., please check if "official national totals" is the correct word

e.g., "… 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." Please revise this sentence. At least, "the impact of L1 flood impact account for ……"

e.g., "CFDID also includes agricultural impact indicators, which are rarely available in international hazard impact databases." Although important, the indicators should be presented along with the other 7 impact indicators.

Response:

We sincerely thank you for these valuable suggestions on the wording of the abstract. We have removed "worldwide", replaced "top-level" with "parent-event", revised the percentage

statement to refer to the impacts reported in L1 records, and introduced all eight impact indicators together.

Regarding the third point, “official national totals” refers here to national flood-impact statistics reported by the responsible Chinese authorities. The “national totals” expression has also been used in scientific publications, such as Zhang et al. (2026).

Reference: Zhang, S., Zhou, L., Liang, H., Obulkasim, O., and Dai, Y.: Underestimated agricultural losses due to flooding, Sci. Adv., 12, eaed2754, https://doi.org/10.1126/sciadv.aed2754, 2026.

The revised abstract is provided below.

Abstract: Lines 9-25:

“China is one of the countries most severely affected by flood disasters. Long-term and systematic flood impact data are essential for flood risk assessment, flood impact model calibration, disaster prevention and reduction. However, an open, long-term, China-focused event-scale database with detailed flood disaster impact indicators remains lacking. Here we present the China Flood Disaster Impact Database (CFDID), which contains 1716 flood disaster event records from 1994 to 2023, including 1391 parent-event (L1) records. CFDID is compiled from 66 official documents and covers five flood-related disaster categories: rainstorm flood, typhoon, snowmelt flood, ice-jam flood, and dam-break flood. Each record includes date, location, disaster category, data source and available values for eight impact indicators: deaths and missing persons, affected population, affected cropland area, damaged cropland area, failed cropland area, collapsed dwellings, damaged dwellings, and direct economic losses. The database was constructed using a workflow that combines optical character recognition (OCR), large language model (LLM)-based information extraction, and manual verification, thereby improving the efficiency of converting unstructured text information into structured event records while ensuring data quality. Relative to the corresponding official national totals during 1994–2023, the impacts reported in 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. CFDID also includes agricultural impact indicators, which are rarely available in international hazard impact databases. CFDID is available at https://doi.org/10.5281/zenodo.22918521 and will be continuously updated by year."

R3C2:

Comment: *L75. The relationship between L1, L2, and L3 is unclear. Introducing a case should help readers understand the necessity of this categorization. And L206, why L2 records are compared with Wikimpacts.*

Response:

We sincerely thank you for this constructive comment. We have added a September 2011 example to Sect. 2.1 to illustrate the nested relationships among L1, L2, and L3 records.

We have also clarified that the comparison uses CFDID L1 records and Wikimpacts L2 records. The level labels are database-specific: CFDID L1 identifies parent-event records, whereas Wikimpacts L2 provides country-level impact information and its L1 records may aggregate impacts across several countries. We therefore use CFDID L1 records to exclude its nested sub-records from aggregation and China-related Wikimpacts L2 records to select the relevant reporting scope. These explanations have been added to Sects. 2.1 and 2.5.

2.1 Definition and classification of flood disaster events: Lines 82-84:

"For example, the L1 record "2011-043" describes the September 2011 flood impacts across six provincial-level regions. Its L2 child record "2011-043-01" reports impacts in Shaanxi. Nested within this L2 record, the L3 record "2011-043-01-01" reports localized impacts in Baqiao District, Xi'an."

2.5 Assessment relative to official annual totals: Lines 215-218:

"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, its L2 national-level records for China are used to select impacts, whereas its L1 records may span several countries."

R3C3:

Comment: *Table 1. "number of sources", should be "number of documents".*

Response:

We sincerely thank you for this comment. We have changed "Number of sources" to "Number of documents" in Table 1.

2.2 Data sources: Lines 116:

"**Table 1. Data sources of CFDID.**

No.	Data source	Year	Number of documents	Responsible authority
1	China Water Resources Yearbook	1994–2018	25	Water resources authorities
2	Yearbook of Meteorological Disasters in China	2002–2023	22	Meteorological authorities
3	Top Ten Natural Disasters in China	2005–2023	19	Civil affairs authorities (2005–2017) / Emergency management authorities (2018–2023)

"

R3C4:

Comment: *Appendix A. Better include an event case rather than only the cover of the books.*

Response:

We sincerely thank you for this constructive suggestion. We have retained the source-document examples in Fig. A1a–c and added an excerpt from the 2023 Yearbook of Meteorological Disasters in China in Fig. A1d. The 2023 Yearbook of Meteorological Disasters reports disaster events that occurred in 2022. The excerpt provides context on heavy rainfall and regional flooding in northeastern China, followed by impact reports for two events in Jilin during

June and July 2022. An English translation is provided below the figure, and Table A1 presents the two corresponding final CFDID records, “2022-002” and “2022-009”. The source page is included in the processing example released with CFDID on Zenodo.

2.2 Data sources: Line 107:

“Examples of the three data sources and an event-description case are shown in Appendix A.”

Appendix A: Examples of data sources: Lines 452-478:



Figure A1. Examples of data sources. (a) China Water Resources Yearbook in 2000. (b) Yearbook of Meteorological Disasters in China in 2003. (c) “Top Ten Natural Disasters in China” in 2023. (d) An excerpt from the 2023 Yearbook of Meteorological Disasters in China (p. 12), which reports disaster events from 2022.

The English translation of the paragraph in Fig. A1d is as follows: “During June–July, mean precipitation across the three northeastern provinces was 334.1 mm, 39% above normal and the second highest for the same period since 1961 (Fig. 2.2.3 in the original source). In Jilin, precipitation reached 414.2 mm, 65% above normal, and the number of rainy days reached 37.8; both were record highs for the same period. In Liaoning, precipitation reached 420.6 mm, 70% above normal and exceeding the normal total for the entire summer. This was the highest precipitation for the same period in nearly 30 years, while the number of rainy days was the second highest since 1961. Heavy rainfall caused water levels to exceed warning thresholds on 40 rivers in the Songhua and Liao river basins. In early August, a dike breach occurred along the Panjin reach of the Raoyang River in Liaoning. Road infrastructure was damaged or disrupted in some areas. Temporary waterlogging occurred in low-lying farmland in parts of Jilin and Liaoning. Together with reduced sunshine, this adversely affected the growth of single-season rice, spring maize, soybeans, and other crops.

Jilin: During 27–29 June, 27,000 people were affected across 12 county-level administrative areas in eight cities, including Siping, Liaoyuan, Baishan, Songyuan, and Jilin City. Affected cropland totalled 5,500 ha, and direct economic losses exceeded CNY 26 million. During 13–16 July, 26,000 people were affected across 10 county-level administrative areas in six cities, including Changchun, Songyuan, Liaoyuan, Siping, and Jilin City. Affected cropland totalled 6,500 ha, including over 200 ha with crop failure, and direct economic losses exceeded CNY 28 million.”

The two date ranges in this paragraph are represented by separate L1 records, “2022-002” and “2022-009”, in the released database (Table A1).

Table A1. Final CFDID records 2022-002 and 2022-009 corresponding to the source paragraph in Fig. A1d.

Event level	L1	L1
Event ID	2022-002	2022-009
Start year	2022	2022
Start month	6	7
Start day	27	13
End year	2022	2022

End month	6	7
End day	29	16
Province	Jilin	Jilin
Specific location	Siping City, Liaoyuan City, Baishan City, Songyuan City, Jilin City, and other affected areas, totaling 8 cities and 12 counties	Changchun City, Songyuan City, Liaoyuan City, Siping City, Jilin City, and other affected areas, totaling 6 cities and 10 counties
Disaster category	rainstorm flood	rainstorm flood
Typhoon information		
Deaths and missing persons (persons)		
Affected population (thousand persons)	27.000	26.000
Affected cropland area (thousand ha)	5.500	6.500
Damaged cropland area (thousand ha)		
Failed cropland area (thousand ha)		0.200
Collapsed dwellings (thousand rooms)		
Damaged dwellings (thousand rooms)		
Direct economic losses (million CNY)	26.000	28.000
Direct economic losses (2023 constant prices, million CNY)	26.052	28.056
Data source	Yearbook of Meteorological Disasters in China	Yearbook of Meteorological Disasters in China

”

R3C5:

Comment: *L194. I don't understand "within-source checking and cross-source fusion", please elaborate on this method.*

Response:

We sincerely thank you for identifying this unclear description. We have revised Sect. 2.4.4 to distinguish within-source checking, which examines repeated descriptions and potential parent–child relationships within each source document, from cross-source fusion, which compares records across different source series. We also specify the candidate-screening conditions and explain the subsequent LLM-assisted relationship assessment and manual verification.

2.4.4 Data fusion: Lines 186-196:

"Duplicate, overlapping, and nested records may exist both within the same source and across the China Water Resources Yearbook, the Yearbook of Meteorological Disasters in China, and the "Top Ten Natural Disasters in China". We therefore conducted data fusion in two stages, including within-source checking and cross-source fusion. First, within-source checking examined records extracted from each source document to identify repeated descriptions and potential

parent–child relationships. Second, cross-source fusion compared records from different source documents to integrate complementary information and identify duplicate, overlapping or nested records.

For both within-source checking and cross-source fusion, candidate record pairs were first identified using three conditions: (1) the same start year, (2) start months differing by no more than one month, and (3) at least one shared province. Pairs not meeting all three conditions were not selected for the candidate comparison. Then the "deepseek-v4-pro" LLM compared the candidate records and their original source texts and classified their relationships as the same event, different events, nested records, or uncertain relationships. The first author manually reviewed all resulting assessments against the source texts."

R3C6:

Comment: *L249. It is unusual to see Fig. 7 between Fig. 2 and Fig. 3 when Fig 7 is still in the main text (rather than supplementary).*

Response:

Thank you for pointing this out. We have removed the early reference to the former Fig. 7 from the section 3.2 and introduce the comparison figure in the detailed EM-DAT comparison.

R3C7:

Comment: *Please reorganize the discussion about the validation in section 4.2, as first please present how good the current EM-DAT data are captured in the CFDID (in terms of event number, impact), and then represent the correlation of the two sources in temporal scale, and finally present how much EM-DAT underrepresents the events that captured CFDID..*

Response:

We sincerely thank you for this constructive suggestion. We have expanded Sect. 3.7 to quantify interannual correlations, record correspondence, and complementary reporting. We retain the annual comparison first to establish the overall temporal context and follow the panel order in Fig. 7, followed by record correspondence and records without identified counterparts.

We now report Pearson correlation coefficients of 0.933, 0.661, and 0.863 between the annual CFDID and EM-DAT series for deaths and missing persons, affected population, and direct economic losses, respectively. For record correspondence, 313 of the 433 (72.29%) EM-DAT records have identified CFDID counterparts, involving 569 CFDID L1 records (40.91%). We then quantify the records without identified counterparts in both databases and their reported impacts relative to official national totals.

3.7 Detailed comparison with EM-DAT: Lines 346-348:

"The Pearson correlation coefficients between the CFDID and EM-DAT annual series during 1994–2023 are 0.933, 0.661, and 0.863 for deaths and missing persons, affected population, and direct economic losses, respectively."

3.7 Detailed comparison with EM-DAT: Line 354:

"In total, 313 EM-DAT records (72.29%) have identified CFDID counterparts, involving 569 CFDID records (40.91%)."

3.7 Detailed comparison with EM-DAT: Lines 356-357:

"Overall, 822 CFDID records (59.09%) have no identified EM-DAT counterpart."

Reviewer 4

R4C0:

General comments: *While abundant flood discharge datasets are available, this paper introduces a dataset of flood disaster impacts across China. A total of 1716 flood disaster events spanning 1994–2023 were compiled via optical character recognition (OCR), large language model (LLM)-driven information extraction, and subsequent manual verification. Overall, the newly constructed dataset is potentially valuable.*

Response:

We sincerely thank you for recognizing the potential value of CFDID and for your constructive suggestions. Your comments have helped us clarify the database's contribution and improve the presentation of its data sources, construction workflow, and comparison results. The main revisions are summarized below:

1. Comparison with existing databases. We have explained the relevance of HANZE and Groundsource to CFDID, and expanded the comparison with EM-DAT to present bidirectional record correspondence and complementary reporting (R4C1–R4C2).

2. Clarification of data sources. We have explained the different reporting scopes of the three sources used in CFDID (R4C3).

3. Workflow presentation. We have substantially revised Fig. 1 to show the processing operations and intermediate outputs, including candidate screening, manual verification and decisions, event ID assignment, and adjustment of economic losses to 2023 prices (R4C4).

4. Manuscript organization. We have moved the database comparisons to the Results section as Sects. 3.6 and 3.7, while retaining potential applications and limitations in the Discussion as Sects. 4.1 and 4.2 (R4C5).

We sincerely thank you again for your careful review and constructive suggestions. Detailed, point-by-point responses and the corresponding manuscript revisions are provided below.

R4C1:

Comment: *Several valuable flood impact datasets have already been published. Representative examples include global datasets such as the DFO Flood Observatory and Google Groundsource, as well as the European HANZE (Historical Floods with Impact Data) dataset. The authors are recommended to prepare a comparison table benchmarking their dataset against existing products to clearly highlight its strengths and limitations.*

Response:

We sincerely thank you for this constructive suggestion. We have clarified the scope of the quantitative comparison table and expanded the accompanying text to discuss HANZE and Groundsource. The table focuses on China-related event records and reported impacts relative to official national totals.

HANZE was an important methodological reference for the field design of CFDID. We have added this explanation to Sect. 2.3, highlighting their shared information categories, including event dates, locations, human and economic impacts, and data sources. HANZE covers Europe and therefore is not included in the quantitative comparison for China.

We also examined the Groundsource preprint and its publicly released v1 dataset. Groundsource provides news-derived information on flood dates and reported locations, but does not include the quantitative socioeconomic impact indicators required for our national-total comparison. We have therefore discussed it in the “Limitations and future improvements” section as a potential resource for identifying additional candidate events and supporting future news-based extensions of CFDID.

These revisions clarify the complementary roles of the datasets and distinguish methodological references and potential future resources from databases suitable for the present quantitative comparison.

2.3 Database fields: Lines 127-130:

"The European flood-impact database HANZE (Paprotny et al., 2024) provided a methodological reference for the field design of CFDID, particularly the recording of event dates, locations, human and economic impacts, and data sources. CFDID's fields reflect the information available in Chinese official sources, including detailed agricultural and dwelling impact indicators."

3.6 Overall comparison with international flood impact databases: Lines 317-340:

"We compare CFDID with three international flood impact databases in terms of event record counts and reported impacts as percentages of official national totals (Table 4). To ensure consistency with the event hierarchy used in Wikimpacts, the comparison uses L1 records for CFDID and L2 national-level China records for Wikimpacts. 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.

During 1994–2023, CFDID contains 1391 L1 event records, exceeding the number of flood records for the Chinese mainland in EM-DAT, DFO, and Wikimpacts. However, record counts are influenced by both the information availability and differences in event delineation and aggregation, and should not be interpreted as the only measure of database completeness.

For the five core impact indicators, CFDID accounts for higher percentages of the corresponding official totals for affected population, collapsed dwellings, and direct economic losses than the international databases. In addition, CFDID provides agricultural impact indicators,

including affected cropland area, damaged cropland area, and failed cropland area, which are rarely available in global databases. These indicators are particularly useful for evaluating agricultural flood impact models (Zhang et al., 2026).

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%

”

3.7 Detailed comparison with EM-DAT: Lines 342-373:

“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 therefore 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. The Pearson correlation coefficients between the CFDID and EM-DAT annual series during 1994–2023 are 0.933, 0.661, and 0.863 for deaths and missing persons, affected

population, and direct economic losses, respectively. 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. In total, 313 EM-DAT records (72.29%) have identified CFDID counterparts, involving 569 CFDID records (40.91%). 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. Overall, 822 CFDID records (59.09%) have no identified EM-DAT 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.

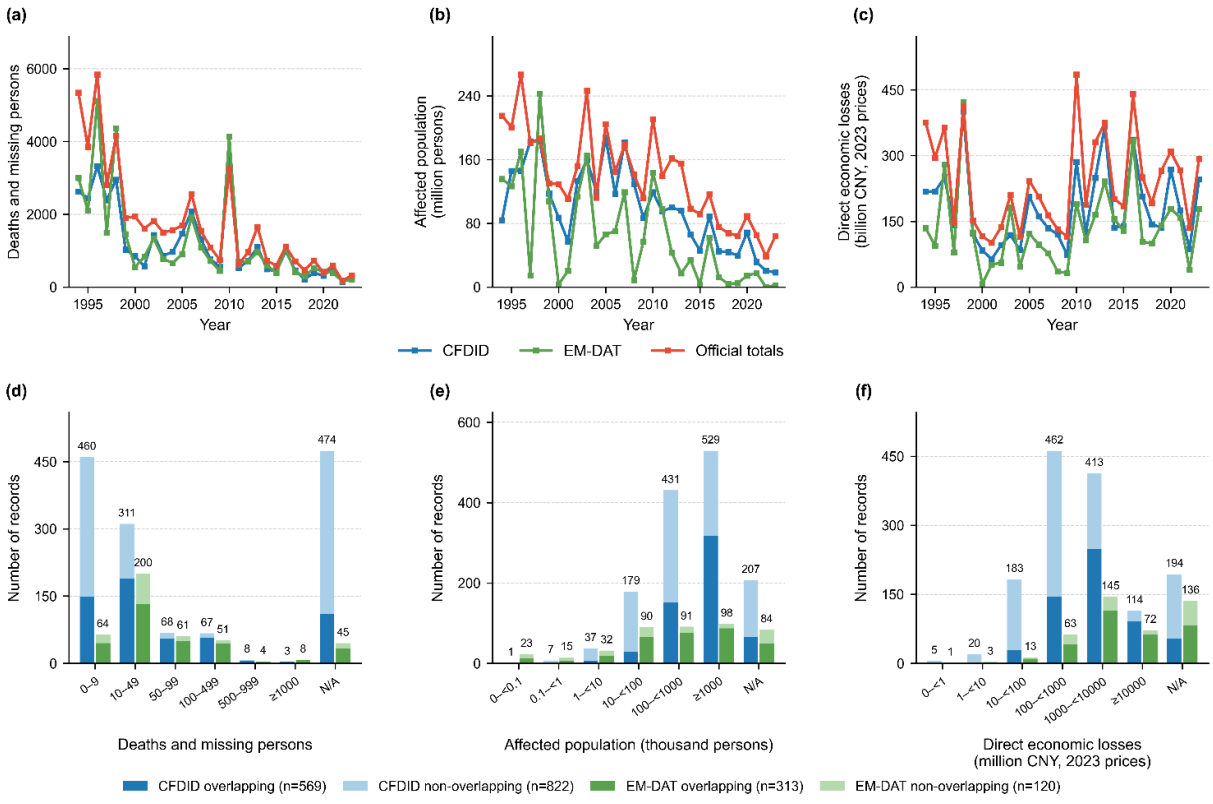


Figure 7. Comparison of flood impacts in CFDID, EM-DAT, and official national totals during 1994–2023. (a–c) Annual deaths and missing persons, affected population, and direct economic losses. (d–f) Record counts by magnitude, with darker and lighter segments indicating records with and without identified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

4.2 Limitations and future improvements: Lines 397-401:

"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. Groundsource v1 also provides a global inventory of flood records extracted from news reports (Mayo et al., 2026). These news- and social-media-based inventories could help identify additional candidate events for future extensions of CFDID, although they do not systematically provide the full set of impact indicators included in CFDID."

R4C2:

Comment: *Beyond the tabular comparison, direct record-level comparison is also essential. Given that datasets often exhibit incomplete spatial and temporal overlap, it would be informative to identify the unique information captured by the newly developed dataset.*

Response:

We sincerely thank you for this constructive comment. We have expanded Sect. 3.7 to present a bidirectional record-level comparison between CFDID and EM-DAT using the matching procedure described in Appendix C. The comparison includes 1,391 CFDID L1 records and 433 EM-DAT records. We identified 822 CFDID records (59.09%) without an identified EM-DAT counterpart. Figure 7d–f distinguishes records with and without identified counterparts across different impact magnitudes. The revised text also reports the impacts documented by the unmatched records relative to official national totals and examines CFDID records meeting EM-DAT's numerical thresholds. The identifiers of matched and unmatched records are available with the CFDID release on Zenodo.

3.7 Detailed comparison with EM-DAT: Lines 342-373:

"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 therefore 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. The Pearson correlation coefficients between the CFDID and EM-DAT annual series during 1994–2023 are 0.933, 0.661, and 0.863 for deaths and missing persons, affected population, and direct economic losses, respectively. 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. In total, 313 EM-DAT records (72.29%) have identified CFDID counterparts, involving 569 CFDID records (40.91%). 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. Overall, 822 CFDID records (59.09%) have no identified EM-DAT 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.

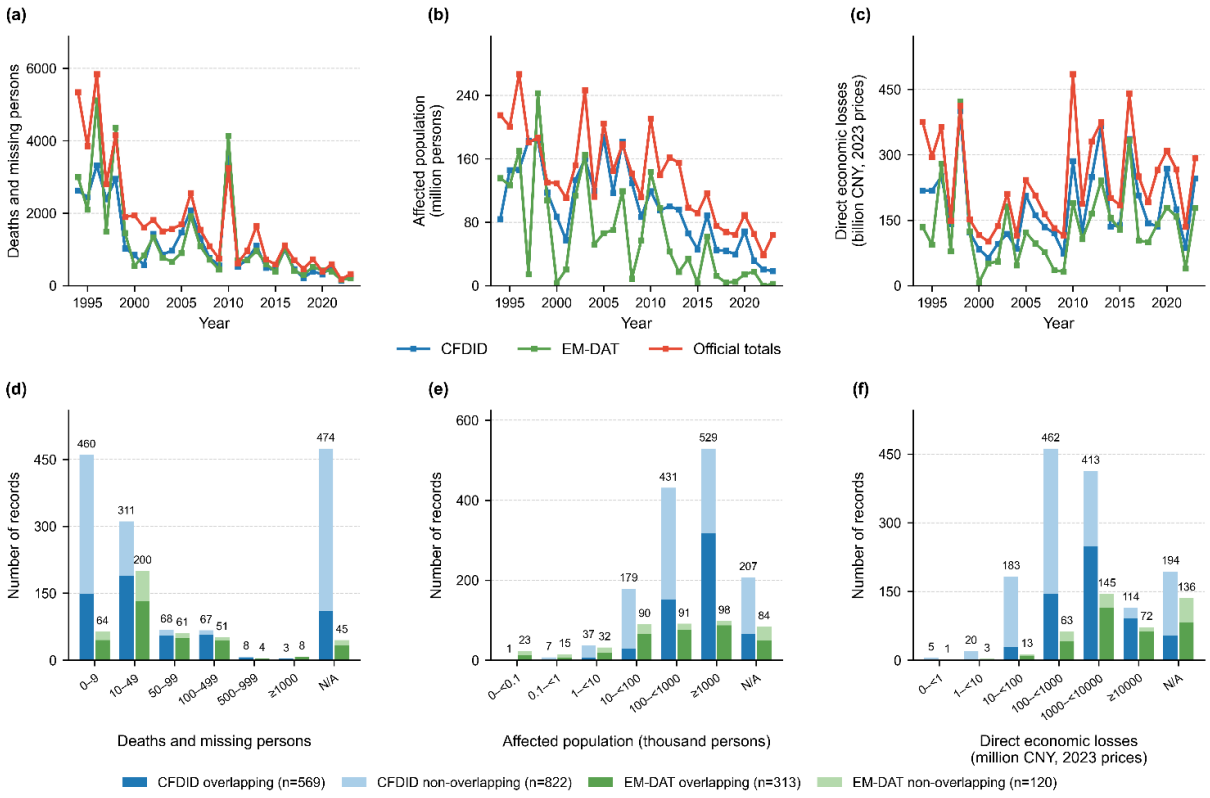


Figure 7. Comparison of flood impacts in CFDID, EM-DAT, and official national totals during 1994–2023. (a–c) Annual deaths and missing persons, affected population, and direct economic losses. (d–f) Record counts by magnitude, with darker and lighter segments indicating records with and without identified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

R4C3:

Comment: *The three data sources listed in Table 1 lack consistency. For instance, the "Top Ten Natural Disasters in China" only documents catastrophic events. The authors should clarify the data sources used for moderate and minor flood events.*

Response:

We sincerely thank you for this constructive comment. We have revised Sect. 2.2 to clarify the different reporting scopes and complementary roles of the three sources. Moderate and minor flood events represented in CFDID are mainly obtained from the China Water Resources Yearbook and the Yearbook of Meteorological Disasters in China, which also document major disasters. The annual "Top Ten Natural Disasters in China" list selects high-impact natural disasters, and only its flood-related entries were included in CFDID. This source therefore primarily supplements information on major flood events.

2.2 Data sources: Lines 114-115:

"The two yearbooks document many flood events of varying magnitudes and the flood-related entries from the annual "Top Ten Natural Disasters in China" mainly list major-event information."

R4C4:

Comment: *Figure 1 appears to be AI-generated. It remains overly conceptual and provides limited substantive information; removal of this figure is suggested.*

Response:

We sincerely thank you for this constructive comment. The original figure was drawn by the authors in PowerPoint and was not AI-generated. We agree, however, that it was overly schematic and provided insufficient detail about the database construction process.

We have therefore substantially revised Figure 1. The revised figure distinguishes the preprocessing of scanned yearbooks from the collection of web-based reports, reports the numbers of text blocks and records at key processing stages, and specifies the different purposes of manual verification after extraction and during record fusion. It also presents the candidate-record screening criteria and clarifies the post-processing operations, including event ID assignment and economic-loss adjustment to 2023 prices.

We have retained the revised figure because it now provides a concise overview of the processing steps and intermediate outputs described in Sect. 2.4.

2.4 Database construction workflow: Lines 151:

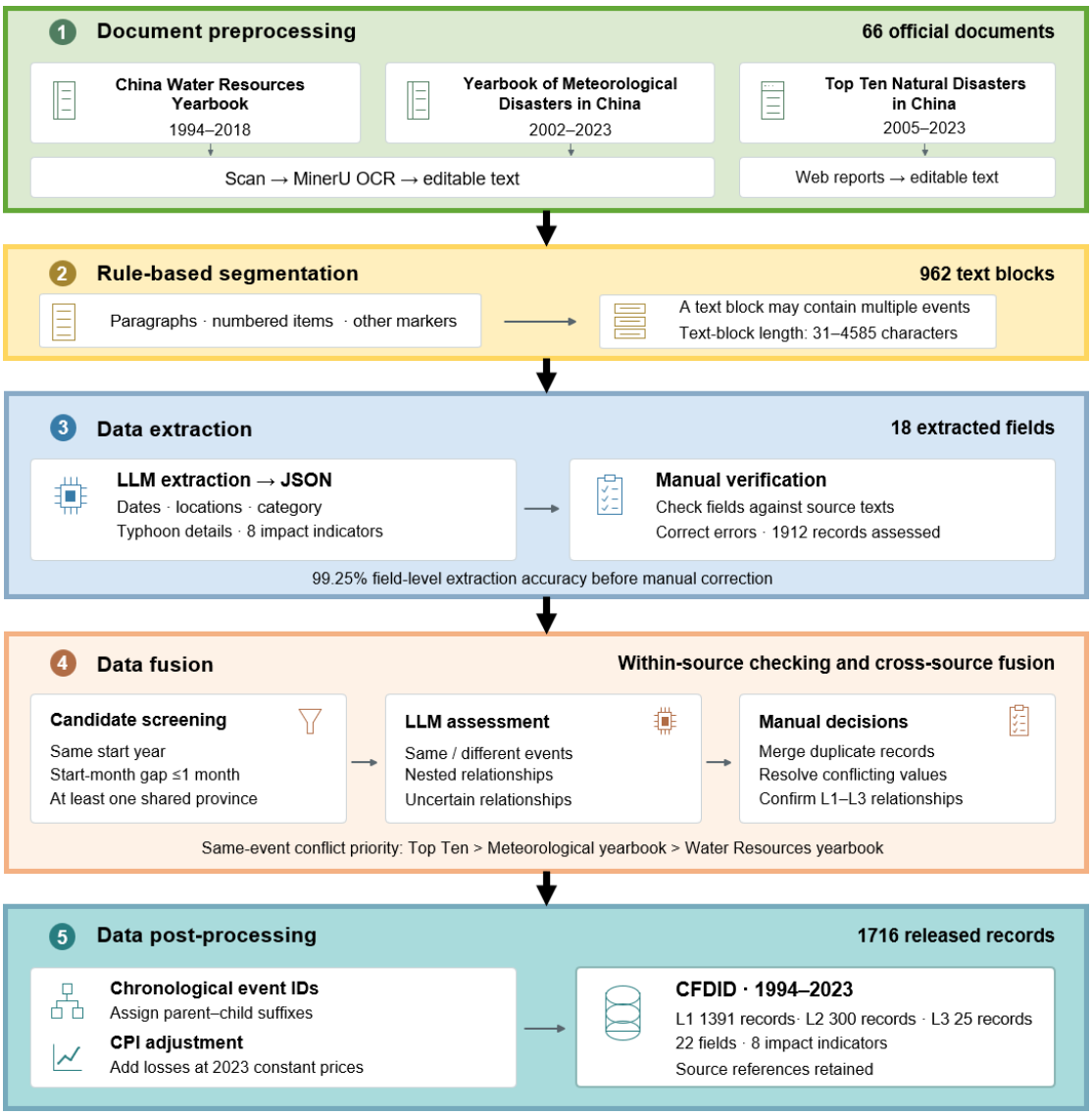


Figure 1. The workflow for constructing CFDID.

R4C5:

Comment: *The current Discussion section contains comparison results against existing datasets. The authors are advised to move the content in Sections 4.1 and 4.2 to the Results section, whereas Sections 4.3 and 4.4 are more suitable for retention within Discussion.*

Response:

We sincerely thank you for this constructive suggestion. We agree that the comparisons with existing databases present key results of this study. We have therefore moved the former Sects. 4.1 and 4.2 to the Results section, where they now appear as Sect. 3.6 (“Overall comparison with international flood impact databases”) and Sect. 3.7 (“Detailed comparison with EM-DAT”).

The former Sects. 4.3 and 4.4 have been retained in the Discussion and renumbered as Sect. 4.1 (“Potential applications”) and Sect. 4.2 (“Limitations and future improvements”), respectively.