

REPLY TO REFREE COMMENTS

RC1: Anonymous Referee #1

“This is a useful study on an important topic. The inclusion of levees in flood extent calculations is a substantial contribution to the flood modelling scientific literature. Overall I find the manuscript well written, the method clearly explained and the test cases useful.”

Reply: Thank you for the positive assessment of the manuscript and for recognising the contribution of explicitly incorporating levees into large-scale flood modelling.

“My biggest concern is that there is not much discussion of the uncertainties coming from the various components of the modelling chain and the potential impacts on the test results presented, as well as the future projections and wider applicability globally. I think this needs some further thought and discussion. Ideally some further sensitivity studies would also be included to demonstrate the impacts of the data/model methodology uncertainties.”

Reply:

We thank the reviewer for this important comment. We have revised the manuscript to more explicitly discuss uncertainties associated with the different components of the modelling chain, including runoff forcing, levee detection and parameterisation, frequency analysis, static levee representation, and volumetric downscaling, and to clarify how these uncertainties influence both event-based validation and return-period results. These points are now integrated into the relevant Discussion sections and synthesised in an expanded Limitations section. In addition, we conducted a targeted sensitivity analysis on the choice of extreme-value distribution and propagated its effect through volumetric downscaling to inundation extent (supplementary Fig. S14), to illustrate how frequency-analysis uncertainty influences high-return-period hazard estimates. We also clarify the implications of these uncertainties for global applicability. While systematic sensitivity experiments would provide additional insight, they would require large ensembles of long-term simulations and are therefore beyond the scope of this study; such analyses are identified as an important direction for future work. The purpose of this expanded discussion is not to fully quantify uncertainty, but to guide readers in interpreting the robustness and limitations of the reported improvements and to avoid overconfidence in return-period projections.

Changes made in manuscript:

Section 4, lines 485 - 498

Section 3.3, lines 424 - 431

Figure S14 (supplementary document)

“The literature review of current global flood models is very thin and a paragraph reviewing this model and methodology in the context of the wider scientific contributions to global flood modelling is necessary, both at the beginning and in the conclusions - detailing how this contribution has wider significance.”

We thank the reviewer for this comment. In the Introduction, we have added a new paragraph that places our work within the wider global flood hazard and risk modelling literature and briefly describes existing global frameworks. In the Conclusions, we have added a short paragraph

explaining that our framework provides a way to convert coarse-resolution outputs from a global river model (CaMa-Flood) into high-resolution, levee-aware hazard maps by using a global levee dataset, and that the same workflow can be applied in other large, leveed river systems. These additions clarify how the study connects to existing global flood models and what its broader significance is.

Changes made in manuscript:

Section 1, lines 49 - 53

section 5 lines 510 -516

“On line 133 there is mention of a group literature review. What is this? Please provide many more details and make clearer the findings and how they fit in this with work.”

Reply:

In the revised manuscript we clarify how the basin-specific design return periods were obtained and used. We now explain that we collected river protection standards for 189 nationally and prefecturally managed rivers by reviewing river maintenance plans and design standard documents published by MLIT and the relevant prefectural governments, and that smaller rivers without such information were excluded. The reported protection levels were categorised by return period (e.g. 50-year, 100-year, 200-year), and one representative design return period was assigned to each river basin, mapped to the corresponding CaMa-Flood basins, and used as the target protection level in the iterative estimation of levee heights. This makes the underlying data and their role in the modelling framework explicit. This clarification is intended to make the origin and assumptions behind the design return periods transparent, so that readers can judge their suitability and limitations when transferring the method to other regions.

Changes made in manuscript:

Section 2.3, lines 169 - 172

“What type of river is the Chikuma River? What are its characteristics and the driving processes of floods and the catchment/floodplain characteristics? Please provide enough information to aid the unfamiliar reader.”

Reply:

In the revised manuscript, we have added a short description of the river in the Methods section. The new text explains where Chikuma River originates, its catchment characteristics and the driving flood processes with proper citations. This contextual information is provided to help readers understand why the Chikuma River represents a suitable test case for levee-dominated flood processes rather than an extreme or atypical basin.

Changes made in manuscript:

Section 2, para 2

“Some further information on what type of model CaMaFlood is, and its major components and applications would be useful in helping this manuscript be standalone. The same goes for the Khanh et al dataset.”

Reply:

In Section 2.2 we have expanded the description of CaMa-Flood model. In Section 2.3 (Levee data), we have expanded the description of the global levee dataset of Khanh et al. (2025, in review) to clarify that it is derived from 10 m LiDAR DEMs using a morphometric detection algorithm based on four terrain parameters. We now also state explicitly that, after resampling to 1-arcsecond resolution, this dataset is used to compute the levee parameters used in our schematisation and volumetric downscaling. These additions make the manuscript more self-contained for readers unfamiliar with CaMa-Flood and the Khanh et al. dataset.

Changes made in manuscript:

Section 2.2, lines 117- 122

Section 2.3, lines 142- 148

“Please check all references appear in the reference list - e.g. Tellman is missing”

Reply:

We thank the reviewer for catching this. We have cross-checked all in-text citations and the reference list. The reference to Tellman et al. (2021) has now been added to the bibliography.

Changes made in manuscript:

References, lines 641 - 643

“In section 3.1.1 you provide only the improved IoU and FBI. These are not much higher than the natural values and a more honest discussion would be valuable here making clearer comparison, and also explaining more clearly to the reader how they can interpret these numbers. This is also the case for the following sections.”

Reply:

We thank the reviewer for this comment. Section 3.1.1 has been revised to include a more transparent and more balanced discussion of IoU, FBI, recall, and precision, explaining how these metrics should be interpreted and why improvements are moderate in levee-dominated floodplains. We also clarify that levee integration primarily improves recall and reduces bias rather than producing significant changes in total inundated area, and note that similar considerations apply to subsequent sections. The revision aims to avoid misinterpretation of moderate metric improvements, especially in levee-dominated floodplains where large gains are not expected.

Changes made in manuscript:

Section 3.1.1, Para 2 and para 3 (lines 326 - 331) and Fig 5

Section 3.1.2, Fig 6

“Section 3.4 is very brief. What are all the major rivers in Japan? What are their characteristics? - can you provide the reader with some details? Again there must be a number of key uncertainties in this approach - these should be shown and discussed here.”

Reply:

We thank the reviewer for this comment. We have expanded Section 3.4 to briefly describe the major river systems considered in the national-scale analysis and to provide context on their typical characteristics relevant to flood protection (mountain-to-plain transitions and extensive leveed floodplains). The added text is meant to prevent over-interpretation of the 10–15% reduction as a direct flood-risk reduction, and to clarify it as an aggregated volumetric signal.

We also added a concise discussion of key uncertainties affecting the flood-volume reduction estimate (levee data uncertainty and resampling, levfrc/levhgt estimation, static levee representation without breach, daily runoff forcing, limited sample length for extremes, and omission of urban/pluvial processes).

Changes made in manuscript:

Section 3.4, lines 434 – 440 and para 4

“More should be included on how this method can be applied globally, particularly on larger rivers that do not occur in Japan. What is the range of applicability?”

Reply:

We thank the reviewer for this comment. We have appended Section 3.4 and the second paragraph of the Conclusions to clarify the method's global applicability and scope. The revised text explains that the current implementation is explicitly designed for CaMa-Flood, as it relies on small-unit catchments and levee-fraction parameters to represent protected and unprotected zones. We further clarify that the approach is most suitable for large, leveed rivers and can be transferred conceptually to other models that provide an equivalent reach-based or subbasin-based representation of river geometry and levee confinement.

Changes made in manuscript:

Section 3.4, last paragraph

Section 5, lines 514 -516

“The limitations section would be better integrated into the relevant parts of the discussion and expanded upon more thoroughly.”

Reply:

We thank the reviewer for this suggestion. We have revised the manuscript to better integrate key limitations directly into the relevant Results and Discussion sections, including the interpretation of historic event validation, comparison with city hazard maps, and national-scale flood volume analysis. In addition, the Limitations section has been updated to provide an overview of methodological constraints and priorities for future development.

Changes made in manuscript:

Section 4 (Old limitation section has been moved here and renamed as "Discussion of limitations and uncertainties")

REPLY TO REFREE COMMENTS

RC2: Jonathan Remo

“Overprediction of flood inundation within levee protected floodplains by large-scale to global river models used in flood hazard and risk assessment is an important issue which needs to be corrected. The methodology here provides a useful framework of constraining flood hazard assessments along leveed river segments which employ global or large-scale river inundation models. However, the current manuscript would benefit from technical and editorial clarifications before being considered for publication.”

Reply: We thank the reviewer for their positive assessment and for recognizing the relevance of addressing flood overprediction in levee-protected floodplains using large-scale river models. We appreciate the constructive suggestions and have addressed the clarifications as detailed below.

General Comments:

1. *“The manuscript would benefit from some discussion on which type of rivers and floodplains this framework would be most applicable to. Given the test area for this model is rivers in the country of Japan, the conceptual model which is implicit in this approach is the quantification of relatively confined floodplains (i.e., hundreds of meters to a few km wide). The authors have not demonstrated or explained if their modeling framework would be applicable to levee floodplains which are only semiconfined or unconfined (i.e. 10s to 100s km wide). Some discussion in the introduction or discussion sections of this manuscript on this issue would help the reader better understand the potential scale limitations of the proposed framework.”*

Reply: We clarified the applicability of this framework by linking it to the CaMa-Flood model structure. CaMa-Flood represents overbank flooding as aggregated floodplain storage at the unit-catchment scale. Accordingly, the proposed downscaling approach is most suitable for confined-to-moderately confined floodplains, where flood storage is primarily controlled by topography and levees. We now explicitly state this in the limitations section. This clarification is intended to help readers understand the spatial scale for which the assumptions of aggregated floodplain storage remain valid, rather than to suggest universal applicability.

Changes made in manuscript:

Section 4, lines 499 - 503

2. *“The filling and draining of floodplains commonly generate a hysteresis phenomenon where water levels and consequently flood inundation extents, depths, and presumably volumes for the same discharge are greater on the falling limb than the rising limb of the flood hydrograph. While the authors state they employed the annual maxima volumes in their study, it would be helpful to mention how they ensured “maxima volumes” were achieved.”*

Reply: We thank the reviewer for the comment. Annual maximum flood volumes were extracted directly from the simulated daily flood storage time series by selecting the single largest flood storage value for each simulation grid (unit catchment) within each hydrological year. This ensures that the peak floodplain-filling state is captured, regardless of whether it occurs on the rising or falling limb of the hydrograph. This procedure was described in detail in the Supplementary

Material (Section S1, Frequency Analysis) and is now briefly highlighted in the main manuscript section 2.

Changes made in manuscript:

Section 2, lines 95,96

Specific Comments:

1. *“Line 26 recommend changing high to large and then qualifying what the authors consider moderate (i.e., 10-year) to larger (>100-year) flood events.”*

Reply: We have updated the wording by replacing “high” with “large” and clarified the interpretation by adding representative return periods.

Changes made in manuscript:

Abstract, lines 27, 28

2. *“Line 59 – GCM – acronyms first use, recommend spelling it out so it is clear to the reader what GCM stands for (Global Climate Model [GCM]?).”*

Reply: We have spelled out General Circulation Model (GCM) at its first use in the introduction.

Changes made in manuscript:

Section 1, line 65

3. *“Somewhere in section 2.3 it would be useful to mention why the Khanh et al., 2025 global levee data set was used over a potentially higher resolution national to regional levee dataset.”*

Reply: We added a brief clarification in the first paragraph of section 2.3, explaining why the Khanh et al. (2025) levee dataset was selected, highlighting its spatial consistency, suitability for large-scale modelling, and relevance for potential global application. This choice was made to prioritise methodological consistency and transferability over local optimisation.

Changes made in manuscript:

Section 2.3, lines 145 - 148

4. *“Somewhere near line 135 it may be worth noting that return periods for levee exceedances (over topping) do not equate to return periods for flood stages (levels) because levees are commonly constructed with freeboard as a factor of safety. In the U.S. this “freeboard” is commonly about a meter for levees with a designed exceedance of 1% annual chance flood (100-year flood). Consequently “100-year levees” will often not be overtopped by the flood stages/levels even with a < 0.2% annual chance exceedance probability.”*

Reply: We clarified in section 2.3 that the design return periods used to estimate levee heights represent target protection standards rather than exact overtopping thresholds due to the presence of freeboard. This clarification is intended to prevent readers from directly equating design return periods with overtopping probabilities.

Changes made in manuscript:

Section 2.3, lines 176, 180

5. *“Figure 3. Please check the x-axis units on the storage curve chart because m3 is too small for most floodplain storage volumes.”*

Reply: We have updated the x-axis of Figure 3 to clarify the storage volume scale and improve clarity without changing the underlying data.

Changes made in manuscript:

Fig. 3 (Left panel)

6. *“Figure 4. The figure description could use more robust description of the downscaling methodology shown here.”*

Reply: We have updated the caption of Figure 4 to reflect better the scenarios represented in the schematic.

Changes made in manuscript:

Caption of Fig. 4

7. *“Line 231. Flood control has generally been replaced by flood risk reduction or mitigation. Philosophically, floods cannot be “controlled”; only the risk they pose to communities can be mitigated.”*

Reply: We have revised the wording to avoid the term “flood control” and to better reflect the role of levees in influencing flood inundation patterns.

Changes made in manuscript:

Section 3.1.1, line 293

8. *“In line 105 the authors define the scenarios assessed a with and without levee. Then in line 239 they switch the without levee scenario to “natural scenario”. Natural scenario, case, or simulation is used in lines 246, 247, 287, 297, 304, 332, 376 and the caption for Figure 8. I recommend sticking to without levee scenario, case, or simulation.”*

Reply: We consistently use the term ‘without levee’ to refer to simulations without flood protection, and have replaced all occurrences of ‘natural scenario/case/simulation’ accordingly in the text and also in the figures.

Changes made in manuscript:

Line 303, 314, 326, 369, 373, 383, 388, 396, 549

9. *“Line 288. I recommend using another word instead of significantly reduced as there was no statistical test performed to assess the significance of the reported reduction.”*

Reply: We replaced the phrase “significantly reduced” with “considerably reduced”.

Changes made in manuscript:

Line 378

10. *“Lines 298 to 300. To my understanding, the modeling framework presented in this study does not evaluate structural performance of the levee which would require a geotechnical assessment of the structure. The ability of the levee to limit inundation for large magnitude floods (500- to 1000 Year RI) is a function of how the levee is parameterized in this modeling framework (over topping or no overtopping). If geotechnical levee dynamics and/or a probabilistic breaching model were applied to the levee along the Chikuma River, then a statement of “structural performance” could be made. ”*

Reply: We revised the text to clarify that the results reflect the effect of levee parameterization rather than the structural performance.

Changes made in manuscript:

Section 3.2; Line 391

11. *“It is also unclear what point the authors are trying to make in lines 300 to 301.”*

Reply: We rephrased the sentence to improve its clarity. The message is that the levees matter a lot below/around design RP, and progressively less at very high RPs.

Changes made in manuscript:

Section 3.2; Line 391 - 394

12. *“Lines 303 to 306. If you are going to make a point about a spatial variability of levee “effectiveness” you need to show the figure supporting this statement in the main paper.”*

Reply: We clarified the text to link the representative example shown in Fig. 8 with the additional supporting figures provided in the Supplementary Material, thereby substantiating the discussion of spatial variability while avoiding repetition of similar kind of figures in the main manuscript.

Changes made in manuscript:

Section 3.2, Line 396 - 397

13. *“Line 335. I recommend clarifying that the authors are referring to a reduction in flood volume in this sentence.”*

Reply: The reported “flood attenuation” refers specifically to a “reduction in simulated flood volume”. We rephrased it in the manuscript.

Changes made in manuscript:

Section 3.4, Line 448 - 449

14. *“Line 335-337. It is unclear what point the authors are trying to make in this sentence. This consistent flood attenuation underscores the levee system’s effectiveness in delaying and reducing floodwater accumulation within unprotected regions. There is a substantial amount of research showing levees increase flood levels and the extent of inundation across up stream or neighboring unprotected areas of the floodplain relative to a levee condition. Please clarify the point the authors are trying to make in this sentence.”*

Reply: We clarified that although levee confinement can locally raise water levels (particularly, where the effective flow width is reduced) even when total flood volume is reduced, the analysis presented in section 3.4 focuses on volumetric responses within levee-protected grids and does not imply reduced flood hazard in neighbouring non-levee areas. This clarification is intended to avoid misinterpretation of volumetric flood attenuation as a reduction in flood hazard outside levee-protected areas.

Changes made in manuscript:

Section 3.4, Line 450 – 452

REPLY TO REFREE COMMENTS

RC3: Jacob Schewe

General Comments:

“This is an interesting and well-structured paper, presenting a flood downscaling method for the CaMa-Flood river model with a representation of levees. The paper builds strongly on the paper by Zhao et al. (2025) who presented the levee scheme for CaMa-Flood, including identification of levee data and estimation of the associated model parameters, levee fraction and height. The contribution of the present paper lies in developing a downscaling approach (from the 0.25 degree resolution of the CaMa-Flood hydrodynamic model to finer resolution) that respects the presence and characteristics of levees. It is therefore a useful complementary contribution to enhance the utility of the CaMa-Flood model with levee representation, and potentially more generally, since the volumetric downscaling is a fundamentally different approach than the depth-based downscaling method available before. The paper appears generally sound and worth of publication, but some details should be improved, especially regarding methodological explanations and figures.”

Reply: We thank the reviewer for this positive and constructive assessment of our work. In response to the reviewer’s suggestions, we have revised the manuscript to improve the clarity of the methodological description and strengthen the presentation of key figures.

Detailed comments

Introduction, Methods:

“lines 67ff: It is very useful to list the study’s main goals here. For goals a) and b), these seem very similar to the goals of the related paper by Zhao et al. (2025). Consider clarifying a bit more how the two papers relate, and what are their respective unique contributions.”

Reply: We clarified that goals (a) and (b) build on the levee schematization framework of Zhao et al. (2025), using levee data from Khanh et al. (2025), and are treated as prerequisites. We explicitly state that the main contribution of this study is goal (c), the levee-aware volumetric downscaling and its application to high-resolution flood hazard mapping in Japan. This clarification is intended to help readers distinguish clearly between the foundational levee schematization work and the new downscaling contribution introduced in this study.

Changes made in manuscript:

Section 1; Line 78 -80

“line 120: The term “levee distance” is used here (and a few lines below) but has not been introduced yet, therefore it is unclear what it exactly means at this point.”

Reply: We defined the term “levee distance” at its first occurrence to clarify its meaning.

Changes made in manuscript:

Section 2.3; Line 150 -151

“line 126ff: Here it would be useful to show a schematic illustrating the different parameters. Consider using Fig. 3 for this purpose (parameter names would have to be added there), or adding a figure. Referring (in addition) to relevant figures from Zhao et al. (2025) may also be useful.”

Reply: We clarified that the geometric definition of levee parameters follows Zhao et al. (2025), while their conceptual role in storage partitioning is illustrated schematically later in the Methods section 2.4.1

Changes made in manuscript:

Section 2.3; Line 161 -163

“(1): rivlen is not defined. A question is, is this equation valid/acceptable if the river, levees, and/or unit catchments have a complex shape?”

Reply: We defined rivlen in the manuscript as the CaMa-Flood/CaMa-Map river channel length parameter and clarified that Eq. (1) follows the unit-catchment sub-grid parameterization, which remains applicable for complex river and catchment geometries.

Changes made in manuscript:

Section 2.3; Line 164 -166

“line 133: the reference to “our group conducted...” is a bit unspecific. Is this literature review published, or are its findings discussed in more detail elsewhere? Then please indicate the reference.”

Reply: We revised the text to remove the vague reference to an internal literature review and instead explicitly describe the compilation of basin-specific design return periods from official river maintenance plans and design standard documents published by MLIT and relevant prefectural governments.

Changes made in manuscript:

Section 2.3; Line 169 -173

“line 136: How was the representative return period determined?”

Reply: We clarified that each river basin has a single officially designated design return period in the management plans, which was directly adopted as the representative return period used in this study.

Changes made in manuscript:

Section 2.3; Line 174 -178

“line 137: Frequent should be Frequency.”

Reply: We corrected the typo and replaced the word “frequent” with “frequency”

Changes made in manuscript:

Section 2.3; Line 184

“line 137f: I have a (perhaps naïve) question: If I understand right, river water depth is the water depth above the bottom of the river channel, including any overbank water. This means the relationship between total water volume and river water depth will be quite non-linear, with water depth rising quickly as long as it’s below bank height, and more slowly as soon as there is overflow. Annual maximum river depth statistics can contain values from either “sides” of the storage-depth curve (i.e., flooding and non-flooding cases). What are the implications of this for the frequency analysis, using a Gumbel distribution? Would it not be more stable to perform frequency analysis on total water storage, or discharge, for instance?”

Reply: As clarified in the manuscript, and consistent with Zhao et al. (2025), we apply frequency analysis directly to annual maxima of CaMa-Flood simulated river water depth. Because annual maxima were extracted from a 41-year simulated time series, the resulting depth extremes naturally include both in-channel and overbank (floodplain) conditions, which are consistently reflected in the frequency analysis. This choice was made deliberately because river water depth is directly linked to levee crest heights, making it the most relevant variable for estimating levee exceedance conditions within the modelling framework.

Changes made in manuscript:

Section 2.3; Line 184, 185

“line 144: What is meant by “spatial changes”, and by the “localized peak”?”

Reply: We explained that “spatial changes” refer to unit-catchment-wise percentage differences in design-return-period flood depth between successive iterations, and that the “localized peak” denotes the maximum of difference in successive iterations.

Changes made in manuscript:

Section 2.3; Line 191, 193

“line 145f: How where levee fractions estimated? The text up to here is only about levee heights.”

Reply: Levee fractions are estimated earlier in the Methods section using Eq. (1) and explained in detail. The paragraph here focuses on the iterative adjustment of levee heights, while the levee fraction estimation itself remains unchanged throughout the iterations.

“line 69: What does low-lying mean here? Anything below the levee crest?”

Reply: The manuscript has been revised to remove the term “low-lying” and to explicitly state that storage is allocated to grid cells inside the levee with elevations below the levee crest.

Changes made in manuscript:

Section 2.4; Line 222, 223

“line 175ff: Can it be said whether the volumetric downscaling has advantages over a depth-based downscaling approach also in the absence of levees? I presume that other, perhaps natural topographic features can also be better represented with the volumetric downscaling, but can this be shown? Perhaps it’s outside the scope of the present paper.”

Reply: We thank the reviewer for this insightful comment. An inherent advantage of the volumetric downscaling approach used here is that it preserves flood storage explicitly, rather than inferring inundation solely from water depth. Because the downscaling operates directly on simulated flood volumes, it ensures strict volumetric consistency while naturally accounting for non-linear storage–depth behaviour, both below and above bankfull conditions. In the absence of levees, the method reduces to a terrain-controlled volumetric redistribution over the unit catchment and therefore already reflects natural topographic controls on inundation. A systematic comparison with depth-based downscaling methods is beyond the scope of the present paper, whose primary objective is to demonstrate a levee-integrated extension of a storage-consistent framework. This discussion is included to clarify the conceptual behaviour of the method, rather than to claim superior performance over depth-based approaches in non-leveed settings.

Changes made in manuscript:

Section 2.4; Line 211, 213

“line 189: While Fig. 2 does usefully illustrate the high-resolution topography data, it does not actually show any areas.”

Reply: Figure 2 serves as a conceptual schematic to demonstrate the elevation-sorting procedure described in the text and formalised in Eq. (2), rather than a spatial map of inundated area. The physical areas are represented quantitatively in Eq. (2), while Fig. 2 visualises the sorting logic utilised in the calculation.

“(3): k is not defined. This equation basically adds up horizontal slices. One could also imagine adding up vertical columns, one for each grid cell. Would that make any sense, or why is this solution here preferable? Just out of curiosity.”

Reply: In Eq. (3), the variable k denotes the index of sequential water surface levels, which are determined by sorting high-resolution floodplain elevations within each unit-catchment. Each difference ($h_k - h_{k-1}$) indicates a small increase in water level, and the summation calculates the corresponding volume of stored water by multiplying this increase by the flooded area at that level. While it is possible to sum vertical water columns for each grid cell as an alternative, the horizontal slicing method used here is more intuitive for developing a continuous depth–storage relationship. This approach aligns directly with the physical process of water level rise, ensures a monotonic storage curve, and matches the storage–elevation formulation in CaMa-Flood. Consequently, it is both reliable and computationally efficient for large-scale applications.

“(4): First, eq. (3) provides the finite difference approximation of the integral in eq. (4); then, the curve derived from eq. (4) (or from eq. (3)?) is again interpolated (between which knots?). Consider clarifying which equation is actually used in the algorithm.”

Reply: Equation (3) is the formulation used in the algorithm and represents a discrete finite-difference approximation of the continuous integral expressed in Eq. (4), which is shown for conceptual clarity. The depth–storage relationship is constructed directly from the discrete elevation levels (h_k) derived from the sorted high-resolution topography. Because these elevation levels are closely spaced, the resulting depth–storage curve is already well resolved, and further interpolation is generally minimal. Linear interpolation is only applied when a target storage volume falls between two adjacent discrete levels, in order to estimate the corresponding water surface elevation.

“line 209ff: In this case, would overflow to neighboring catchments occur? Is this accounted for in the model?”

Reply: No overflow to neighboring unit catchments occurs at the downscaling stage. The downscaling procedure operates on flood volumes simulated by CaMa-Flood at the coarse-resolution grid level, and each simulated volume is redistributed strictly within its corresponding high-resolution unit catchment. This is done intentionally to maintain consistency with the CaMa-Flood model formulation, in which lateral exchanges between unit catchments are already resolved during the hydrodynamic simulation. The downscaling step therefore does not introduce any additional cross-catchment flow; instead, it redistributes the simulated storage locally within each unit catchment while preserving the original mass balance.

“line 219: Please explain catmz100.”

Reply: catmz100 is a CaMa-Flood high-resolution map parameter that represents the relative pixel position within a unit catchment, normalized from 0 to 100. It is used here to classify pixels as inside or outside the levee based on the levee fraction.

Changes made in manuscript:

Section 2.4.2; Line 279, 280

“(6): The symbol \sim is unclear/ambiguous. Here it shall probably mean the complement, i.e. the cells that are not inside cells. Consider using more conventional notation e.g., from set theory, or explaining in the text.”

Reply: The ambiguous ' \sim ' symbol in Eq. (6) has been replaced with an explicit inequality to clearly define outside pixels as the complement of inside pixels.

Changes made in manuscript:

Section 2.4.2; Line 282

Results:

“The first paragraph in section 3.1.1, esp. lines 230-236, summarize the achieved improvements; consider moving this text to the end of the results section, or the Conclusions section.”

Reply: We thank the reviewer for the suggestion. We intentionally kept this paragraph at the beginning of Section 3.1.1 to briefly summarise the key findings before presenting the detailed regional results and quantitative comparisons that follow. This helps guide the reader through the subsequent analysis. The main outcomes are also synthesised at a broader level in the Conclusions section, and therefore, we did not move this text.

“line 239: Introduce/define metrics such as IoU and FBI (also, accuracy, in line 241), using equations if necessary.”

Reply: Definitions of the evaluation metrics have been added at their first occurrence in Section 3.1.1. Specifically, brief explanations of Intersection over Union (IoU), Flood Bias Index (FBI), recall, precision, and accuracy are now provided to clarify how model performance is assessed in the flood-extent comparisons.

Changes made in manuscript:

Section 3.1.1; Lines 303, 304

“line 242: “upper half” = northern? Or upstream?”

Reply: The ambiguous term “upper half” has been rephrased with “northern part of the region [b]” to indicate the spatial reference clearly.

Changes made in manuscript:

Section 3.1.1; Line 321

“line 268: Could you briefly explain how these one-in-1000-year rainfall events have been derived?”

Reply: We thank the reviewer for this question. The one-in-1000-year rainfall events referenced here are not derived in this study. They originate from the official City flood hazard maps, which are produced by local and national authorities. These hazard maps are based on design rainfall scenarios defined by the municipality, typically derived from long-term rainfall observations using frequency analysis in accordance with governmental flood-management guidelines.

Changes made in manuscript:

Section 3.1.2; Lines 348, 349

“line 272: In this design level of around 100yr uniform across Japan?”

Reply: Thank you for the clarification request. The reference to a design level of “around 100 years” in this paragraph is specific to the Chikuma River case shown in the example figure, for which the official river protection standard is 100 years. We did not intend to imply that a 100-year

design level is uniform across Japan; design return periods vary by river and management authority. The text has been clarified to make this case-specific.

Changes made in manuscript:

Section 3.1.2; Line 354

“line 283: Is unit catchment A different from region [a]? Are they related? Also, can the levee fraction of 0.39 somehow be recognized in Fig. 7?”

Reply: Region [a] and Unit Catchment A refer to different spatial scales and are related hierarchically. Region [a] denotes a reach-scale validation area used in Section 3.1 and consists of multiple CaMa-Flood unit catchments. In contrast, Unit Catchment A is a single CaMa-Flood unit catchment selected in Section 3.2 to illustrate the fine-scale effects of levee integration on inundation depth and extent. The levee fraction value (e.g., 0.39 for Unit Catchment A) is a model parameter representing the fraction of the unit catchment located inside levees and is therefore not directly visualised in Fig. 7. Instead, Fig. 7 shows the physical levee lines and the resulting inundation patterns. The influence of the levee fraction is reflected implicitly through the reduction in inundated area.

“line 288ff: Would it be possible and meaningful to validate these results using observational data, analogous to Fig. 5 and 6?”

Reply: Figure 7 is intended to illustrate the local effect of levee integration at the unit-catchment scale, rather than to provide an additional observational validation. While observed inundation extents are available at reach and basin scales (Figs. 5 and 6), they are not consistently available at the scale of individual CaMa-Flood unit catchments. Fig. 7 presents a relative comparison between with-levee and without-levee simulations for the same event and location, isolating the impact of levee parameterization while keeping all other factors unchanged. Quantitative validation against observations is therefore performed at the appropriate spatial scales in Figs. 5 and 6. This framing is intended to isolate the methodological effect of levee parameterization rather than to replace observational validation performed at larger spatial scales.

“Line 295ff: In addition to comparing flood areas, would it be relevant to look at the reduction in population exposure (or building/infrastructure exposure) due to levees? Perhaps the gains are even larger there, given levees are may usually be designed to protect a maximum of populations/settlements.”

Reply: We agree that assessing reductions in population or asset exposure due to levees would be valuable. However, this study focuses on developing and evaluating a levee-integrated flood hazard mapping framework, rather than a full risk or exposure analysis. Such assessments require additional population and asset datasets and assumptions, and are therefore left for future work.

“lines 298-302: These two sentences seem to contradict each other: Either “levee protection continues to limit flood spread” even at higher return periods, or “the levee effect for floods with a return period higher than the design return period will be insignificant”. Please clarify.”

Reply: Thank you for the clarification request. The text has been revised to clarify that levees continue to reduce flood extent compared to the without-levee scenario even at high return periods, but their effectiveness diminishes beyond the design return period (~100-year). Therefore, levee effects remain but become gradually weaker. This clarification is intended to prevent interpreting levee effectiveness as binary, and instead to highlight its gradual reduction with increasing event magnitude.

Changes made in manuscript:

Section 3.2; Lines 390 - 394

“line 320: The broadening into urbanized areas is not visible in Fig. 9. Perhaps urban areas can be overlaid in the plots? Also, the nonlinear increase in flood magnitude and extent, and the increases beyond RP100, mentioned in the following sentences, are not visible from Fig. 9 either.”

Reply: The expansion into urbanised areas is intended as a qualitative description based on the background street map, not a quantitative land-use analysis; we have revised the text accordingly to avoid over-interpretation. Regarding the nonlinear increase in flood magnitude and extent, this is reflected in Fig. 9 through the depth-class area summaries shown in the text boxes. These indicate a disproportionate increase in both total flooded area and higher-depth classes between RP 10 and RP 100, supporting the statement of nonlinear growth even though RP > 100 cases are not shown in this figure.

Changes made in manuscript:

Section 3.3; Lines 415 - 416

Conclusions, Limitations:

“line 354f: “Beyond RP 100, the growth in flood extent becomes nonlinear” – again, while this is really interesting, it is not shown.”

Reply: The reference to nonlinear growth beyond RP 100 is based on the progression of inundation area and depth classes discussed in the Results section (Fig. 9 and associated text). We have revised the conclusion to clarify that flood extent increase markedly up to RP 100, with indications of nonlinear behaviour inferred from higher-depth class expansion, rather than presenting it as a directly demonstrated result beyond RP 100. This revision was made to ensure consistency between the strength of the conclusion and the evidence explicitly shown in the Results.

Changes made in manuscript:

Section 5; Lines 525 - 527

“line 366: “bias at higher return periods” – perhaps a reference can be provided for this statement.”

Reply: We have rephrased the sentence and added a supporting reference in the Limitations section to substantiate the statement that estimating high return periods from limited sample lengths can lead to increased uncertainty and potential over- or under-estimation

Changes made in manuscript:

Section 4; Lines 476 - 477

Figures:

“1: Explain what the white/blank areas are. Is the levee fraction zero there, or are these cells masked because they belong to a different river that is not considered here?”

Reply: Levee fractions and heights are computed only for river-adjacent unit catchments where levee features are detected in the levee inventory. White (blank) areas in Fig. 1 indicate unit catchments where no levee data are assigned and which are therefore treated as natural/without-levee cells in the CaMa-Flood simulation. Flooding in these areas follows the natural model dynamics. We have clarified this in the figure caption.

Changes made in manuscript:

Fig 1 caption

Area shown in Fig. 1 seems to only partially overlap with key map in Fig. 5, and no overlap with insets a and b in Fig. 5. Might be useful to see the levee structure in a and b.

Reply: Figure 1 is intended to illustrate the parameterisation step of levee fraction and levee height at the CaMa-Flood simulation grid (1-arcmin) scale, and therefore shows a representative river reach rather than the validation sub-regions shown in Fig. 5. Figure 5, in contrast, presents downscaled inundation results at 1-arcsecond resolution for selected validation areas (insets a and b). Because these figures serve different purposes and operate at different spatial resolutions, a direct spatial overlap is not required. Levee structures influencing regions a and b are implicitly accounted for through the levee parameters derived at the simulation-grid scale and applied during downscaling. Adding explicit levee overlays to Fig. 5 would reduce clarity without adding new information beyond what is already conveyed by the levee-integrated results.

“2 needs a bit more explanation in the caption. Is this for an (illustrative) catchment with six high-resolution grid cells? What is the dashed line? What does flddif stand for? Also I think the figure is not exactly “illustrating the number of grids considered for inundation area calculation at a specific depth increment”, but simply showing the elevation of each grid cell. Consider rephrasing.”

Reply: We revised the caption of Fig. 2 to clarify that it is a schematic example with an illustrative number of high-resolution grid cells, to define the dashed line as an example water level h_k , and to explain flddif as the CaMa-Flood high-resolution elevation-above-channel parameter.

Changes made in manuscript:

Fig 2 caption

“3, line graph: This is slightly confusing because it is not clear whether the vertical axis refers to the water surface level inside or outside the unprotected area. I think the confusion arises because the outside storage is included within the same graph as the inside and total storage. An inside water level of 8.5m corresponds to any outside water level between 7m (approx.; bottom of the

orange line) and 8.5 m, because the outside storage starts filling only once the inside water level has reached the levee height. While the graph seems to suggest that an inside water level of 7m corresponds to an outside water level of 7m. I'm not sure how to best solve this, just pointing out this ambiguity in the figure."

Reply: Thank you for pointing it out. The vertical axis in Fig. 3 represents the water surface level inside the levee (river-side), which governs the hierarchical filling process. As described in Section 2.4 and Section 2.4.2, storage is first filled inside the levee up to the levee crest; only after this threshold is exceeded, outside-levee storage begins to fill. We have clarified this hierarchy explicitly in the figure caption to avoid misinterpretation.

Changes made in manuscript:

Fig 3 caption

"5: In the caption, "neutral" should be "natural", I assume. In the [a] panels, overestimation appears along a narrow strip corresponding to a secondary river (Asa river). Should/could this river be masked? Or simply explain that this is a permanent water body?"

Reply: Thank you for the comment. We confirm that "neutral" was a typographical error and has been replaced with "without-levee" throughout the manuscript to maintain consistency. The narrow band of apparent overestimation in region [a] corresponds to the Asa River, which is a permanent water body. In the observational inundation dataset, permanent water bodies are masked out, whereas in the simulated inundation maps, this narrow river is retained. This mismatch leads to a localised overestimation along the Asa River in the comparison. We have clarified this point in the figure caption. No additional masking was applied to the simulated results to maintain consistency with next figure 6.

Changes made in manuscript:

Fig 5 caption

"6: If "observed inundations" refers to city hazard maps, then do these really represent observations?"

Reply: We agree that the term "observed" was inappropriate in this context. The figure caption has been revised to replace "observed inundations" with "city hazard map extent," clarifying that the comparison is made against a scenario-based reference hazard product rather than direct observations.

Changes made in manuscript:

Fig 6 caption

“7: The order of natural and levee simulations is different here than in Fig. 5 and 6. Consider swapping either here or there, for consistency.”

Reply: Thank you for pointing this out. We have swapped the levee and without-levee panels in Fig. 7 to ensure consistency with Figs. 5 and 6.

Changes made in manuscript:

Fig 7

“9: Subpanels are missing labels for the two regions a and b.”

Reply: Thank you for the comment. We inserted labels for two regions [a] and [b] in the titles of subpanels, making it consistent with the Fig 5

Changes made in manuscript:

Fig 9; title of subpanels

Citations/references:

“On line 45, the reference to Kimura et al. (2022) should probably be corrected to 2023, as there is no 2022 paper by this author in the reference list.”

Reply: Thank you for pointing it out. We have corrected the reference in the text to Kimura et al. (2023)

Changes made in manuscript:

Section1; Line 46

“The reference list includes Mester et al. (2023) which however is not cited in the text. Either include citation if relevant, or remove.”

Reply: Thank you for highlighting it. We removed Mester et al. (2023) from the reference list

Changes made in manuscript:

References; Line 624-625