Introductory statement

We are grateful to the referee for the careful and constructive engagement with our manuscript. The quality and depth of the comments have helped us identify areas for improvement and have contributed substantially to strengthening the manuscript. We have carefully considered each comment and provide detailed responses and argumentation below. We hope that our responses address the referee's concerns adequately.

Referee's comments are numbered and addressed individually and numbered for clarity. The comments are presented in Segoe UI font italic, font size 10.

Responses to the comments are presented in Times New Roman font, font size 11.

Response to the referee's comments

General comments

Comment 1: *The manuscript presents a comprehensive modelling effort aimed at reconstructing the flooding associated with Cyclone Gabrielle over a very large region of New Zealand. The amount of work involved in developing, calibrating and validating the modelling framework is considerable and deserves recognition. I also particularly appreciated the balanced and scientifically honest discussion of the limitations and uncertainties affecting the proposed framework, which considerably strengthens the manuscript. In my opinion, the manuscript is suitable for publication after revision. While the overall quality of the work is good, I believe that several aspects of the modelling framework and its presentation require further clarification before publication.*

Response: We are grateful that the referee recognises the usefulness of the dataset. Many of the comments involve the addition of discussions about modelling choices and implications on simulation results. While some of them are already partly addressed, a restructure of the manuscript is necessary to balance the introduction of the requested discussion points and ensure efficient delivery of information. We are proposing the introduction of a discussion section where most of these points would be addressed. The discussion will directly follow the technical validation and will contain:

1. Uncertainty assessment and model limitations moved in the discussion
    - Includes a discussion on levee breaching methodology – see ***Comment 10***
    - DEM conditioning uncertainties will include a discussion on the impact of ignoring hydraulic infrastructures in waterways on floodplain hydrodynamics – see ***Comment 11***
2. Discussion on general dataset quality
3. Dataset usage and framework applications

Specific comments

1 Background and summary

Comment 2: *This section provides a detailed description of Cyclone Gabrielle and its impacts in New Zealand. However, the scientific context of the study remains relatively limited. The manuscript would benefit from a broader review of recent developments in catchment-scale flood modelling and hydrodynamic simulation of extreme rainfall events, including a clearer discussion of the methodological context of the proposed framework and its relationship with existing regional flood-modelling approaches. This would help readers better understand the methodological contribution of the work beyond the specific New Zealand case study.*

Response: We agree that a dedicated review of the recent regional to national hydrodynamic and flood-modelling approaches would be beneficial to clarify the gap this dataset aims to fill. A paragraph will be included that will present the general scope of global and regional flood modelling framework which typically operate at 30-100 m resolution, the use of hybrid hydrological-hydrodynamics framework and the growing interest for regional flood hazard mapping at resolution exceeding 10 m to inform risk assessment, emergency response planning and enhance preparedness.

3.3 Domain creation

Comment 3: *The proposed framework relies on splitting the catchment into an upper hydrological domain and a lower 2D hydrodynamic domain connected through a large number of river injection points. This introduces a relatively complex set of internal boundary conditions and, according to the manuscript, also requires the use of a 500 m buffer of* ***ghost cells*** *to ensure numerical stability at the injection boundaries. I suggest that the authors provide a clearer discussion of the rationale behind this modelling strategy and its advantages. In particular, Table 3 seems to indicate that, at least for the Waiapu catchment, a different modelling strategy ("no river injections") was adopted. If this interpretation is correct, it would be useful to explain the rationale behind this alternative approach and discuss how it relates to the general framework described in Section 3.*

Response: We apologise for the confusion caused by the inappropriate use of the term "ghost cell" (see ***Comment*** *7*). This has been removed. The hydrodynamic solver runs on GPUs. They offer improved computation performances compared to CPUs due to their built-in parallel computing capabilities, however they usually come as expensive computation units on HPCs and are limited in number. The hybrid hydrological-hydrodynamic modelling strategy was adopted to adapt to these constraints: limit domain sizes and reduce computation requirements. This indeed introduces a large set of river injections to represent river flows entering the domain – over 100 in most domains. The modelling framework was designed as a modular tool precisely to automate the creation of the hydrodynamic solver input and parameter files, from pre-existing files or input produced by parent framework modules, with relatively low processing overhead. Domains' reconstruction and injection point identification, elevation reconstruction, rainfall generation, hydrological modelling and hydrodynamic modelling are managed as 5 separate modules. In the present case, rainfall and hydrological simulations were pre-computed. Given pre-existing input datasets, the cumulated time of the hydrodynamic module pre-processing steps is of the same magnitude as BG-Floods cumulated I/O time and do not exceed 10 min for the largest catchments, that is negligible overhead compared to GPU runtime. This includes topography maps, rainfall maps, hydrographs and tide timeseries extraction, conversion and formatting, creation of the BG-Flood input parameter file and submission to the HPC queue.

Computation savings are substantial. For the Waiapu catchment, computation runtime inflated by ~50% using the catchment-scale model compared to the hydrological-hydrodynamic framework (25h to roughly 40h). This was for a catchment located in the higher end of lower to total catchment surface ratio (close to 0.5), and computation savings scale with larger catchment size and smaller lower to total catchment surface ratio. The framework also enables rapid creation of submodels for rapid scenario testing such as the Esk Valley submodel in the Aropaoanui catchment.

This will be summarised in the framework presentation and potentially extended in the discussion.

3.4 Topography reconstruction

Comment 4: *Section 3.4.1 states that hydraulically conditioned DEMs are generally produced at 8 m resolution, whereas a 4 m DEM was generated for the Heretaunga Plains, where the levee network is more complex. Please clarify the resolution of the original LiDAR-derived topographic dataset and the processing workflow used to obtain both the 8 m and 4 m DEMs. This would help readers understand the effective topographic information content of the simulations.*

Response: The resolution of the raw LiDAR point cloud used for the topography reconstruction is now introduced at the beginning of the section. A short description of the conversion from LiDAR to gridded datasets has been added to complement the description of the hydraulics conditioning. Additional edits were also integrated to clarify the rationale behind the levee conditioning procedure, outline its validation, and explain why 4 m resolution DEM is used in the Heretaunga Plains. The 4 m resolution DEM is constructed with the same LiDAR source as the 8 m resolution and does not introduce new elevation information. Suggested corrections:

> **"**The unstructured raw LiDAR point clouds were first converted into cartesian DEMs. The elevation in each cell of the DEMs was computed as the average value of the LiDAR points elevation values within a disk of half the cell diagonal around the cell centre**"**
>
> **"**The elevation maps resolution is chosen to find the best compromise between resource allocation and accurate representation of the landscape. Overall, 8 m resolution DEMs capture most flow-constraining topographical features. However, levee crest elevations are consistently underestimated where levee widths (~5–15 m, spanning 1 to 3 cells) are of the same order of magnitude as the grid resolution. Comparison with the input 1 m LiDAR shows levee crest elevation differences of up to 0.4 m, which can critically misrepresent flood extents when flood waters are close to overtopping the levee. Consequently, targeted levee conditioning was applied to the DEM in relevant domains. For each cell representing a levees crest, the corresponding LiDAR points in the levees cross-section were extracted and the maximum elevation was used as new cell value.
>
> Levees conditioning was applied in the Heretaunga Plains and Gisborne DEMs, which together account for over 90% of the regions' total levees in length (166 km in the Heretaunga Plains; 87 km in Gisborne) according to the New Zealand Inventory of Stopbanks (NZ idiom for "levees") (NZIS) (Crawford-Flett et al., 2022). In these areas, the levees serve primarily as flood protection infrastructures for the main urban centres and high-value arable land (orchards and vineyards) bordering the rivers. In comparison, Mahia beach, the domain with the third longest levee network, has only 5 km of levees which are mostly confined to the upper reaches of the simulation domain.
>
> The levee conditioning approach was validated in the Heretaunga Plains where the levee network is most complex. Hydrodynamics simulations were compared using the levee conditioned 8 m DEM and a 4 m DEM purposefully created following the same GeoFabrics procedure. Both simulations produced comparable flood extent with maximum water depth differences not exceeding ±0.1 m.**"**

Comment 5: *Lines 180–181. The statement that "a surface roughness of 4 m across the river channel was chosen" is difficult to interpret. Please clarify the definition, units and physical meaning of this parameter, and explain how it relates to the estimated obstruction caused by debris accumulation at the bridge piers.*

Response: The surface roughness length is a characteristic vertical length scale of unresolved or subgrid scale topographic features used by the Smart friction model to estimate the surface friction through a rough-wall formulation. It is expressed in metres and is derived from information on subgrid-scale topographic variability derived from LiDAR statistics at each cell and land cover data. "Surface roughness is replaced by "surface roughness length" for clarity and its physical meaning is introduced upon first encounter as follow

> "Topographic data are represented as gridded digital elevation models (DEMs), which contain the direct surface elevation at given resolution, and surface roughness length maps, which represent the vertical scale of unresolved topographic features used to estimate the surface friction using a rough-wall formulation (Smart, 2018)."
>
> At the bridge piers, the debris accumulated act as an obstruction that locally slows down the flow. The assigned roughness length of 4 m to represent the debris at the bridge piers therefore should be interpreted as an equivalent hydraulic roughness representing the additional drag introduced by the debris, rather than as a direct measure of debris height. The description is updated in the manuscript as follows.
>
> "Finally, in the Heretaunga Plains, large amounts of woody debris and sediment accumulated at the piers of the Brookfield bridge (Tūtaekurī River) were found to locally exacerbate overtopping and runoff directly upstream from the bridge. Sensitivity runs showed that the accumulation is best represented through local increase of the surface roughness rather than increase of the bed elevation in the DEM (i.e as an impermeable physical obstruction equivalent to a dam). A static uniform roughness length of 4 m assigned to the bridge cells across its section provided a satisfactory representation of the additional flow resistance associated with the accumulated debris at the peak of the event."

3.5 Gridded rainfall

Comment 6: *Reading Section 6.1.2 (Hydraulic forcing), it appears that the New Zealand Convective Scale Model (NZCSM) forecast is used to guide the interpolation of rain-gauge observations. However, the nature of the final VCSN rainfall product is not entirely clear from the description in Section 3.5 alone. Please clarify whether it should be interpreted as an observation-based rainfall reconstruction constrained by the forecast, or as a forecast-based rainfall estimate corrected using observations. A clearer description of the respective roles of observations and model forecasts in the interpolation procedure would improve the understanding of the rainfall forcing used in the simulations.*

Response: In the present case, choosing between the two interpretations is not straightforward. In a general sense, the VCSN rainfall product is designed as an observation-based rainfall reconstruction guided by synoptic forecasts: rainfall gauge records are the primary forcing data and the NZCSM forecast constrains the interpolation between the gauge points to ensure realistic rainfall patterns at ungauged locations. In data-dense areas (e.g. Heretaunga Plains), this interpretation is effectively accurate as the density of gauges limits the propagation of NZCSM forecast errors into the interpolated dataset. The interpretation becomes more nuanced in data-sparse regions though (e.g. Waiapu). There, the NZCSM forecast carries greater weight in the interpolation, and NZCSM forecast errors propagate more directly into the interpolated dataset, consistent with the rainfall uncertainty discussion. In this configuration, the VCSN rainfall product lies in-between an observations-based estimate guided by the forecast and a forecast-based estimate corrected using observations. Users of the dataset in gauge-sparse catchments should therefore interpret the rainfall forcing as carrying larger uncertainty than in well-gauged areas, consistent with Table 2. Proposed correction:

> "The hourly rainfall data for the period 2023-02-11 22:00:00 to 2023-02-16-21:00:00 was estimated across the Hawke's Bay and Tairāwhiti-Gisborne regions on a Virtual Climate Station Network (VCSN) (Tait et al., 2006) with a resolution of 0.005°×0.005° in WGS84 [EPSG:4326] (~490 m×490 m in local NZTM projection [EPSG:2193]). Collated rainfall observations from gauges operated by ESNZ, the national meteorological agency MetService, and the HBRC and GDC regional authorities (Fig. 1*a*) are interpolated onto the grid using a thin-plate smoothing spline interpolation, with rainfall forecast of the event generated using the New Zealand Convective Scale Model (NZCSM) (Revell et al., 2019) on a 1.5 km×1.5 km resolution grid as spatial co-variate. In this way, realistic rainfall patterns calibrated by rain gauge observations capture patterns of where the highest rainfall is expected. Observations from citizen rain gauges collated by Hawkes Bay Regional Council are included in a second iteration of the interpolation.
>
> In data-dense areas, the interpolation is dominated by the observations, and the density of gauges limits the propagation of NZCSM forecast errors into the interpolated field limiting the degree of uncertainty of the rainfall input. In data-sparse regions however, the NZCSM forecast assumes a larger portion of the interpolation and NZCSM forecast errors propagate more directly into the interpolated field, increasing the level of uncertainty. Therefore, users of the dataset in gauge-sparse catchments should interpret the rainfall forcing as carrying larger uncertainty than in well-gauged areas, consistent with the classification in Table 2."

3.7.2 Boundary conditions

Comment 7: *Regarding the treatment of the inland boundaries, if I understand correctly, discharge hydrographs are prescribed at the river injection points, whereas the remainder of the upstream boundary is treated as an impermeable wall. The latter boundary condition appears to be applied along the outer edge of the 500 m buffer zone, rather than directly along the lower catchment boundary, in order to prevent numerical instabilities and artificial water accumulation from affecting the solution within the area of interest. Please confirm whether this interpretation is correct and clarify the role of the buffer zone and ghost cells in the implementation of the upstream boundary conditions. Such a discussion would also help readers better appreciate the advantages and limitations of the adopted coupled hydrological– hydrodynamic framework compared with more integrated catchment-scale modelling approaches.*

Response: We understand the confusion. Rainfall forcing is imposed as rain-on-grid within the lower catchment and the injection points are located at the boundary of the lower catchment. In BG-Flood, hydrological forcings (discharge hydrographs) are not treated as prescribed boundary fluxes, but as local areal source terms imposed over a prescribed square surface within the simulation domain. To ensure that both rainfall and hydrological forcings all occur well within the boundary of the simulation domain, the simulation domain is constructed as the lower catchment plus a 500 m buffer zone around it. The term ghost cell was incorrect here and has been removed from the manuscript. The section will be updated to detail the implementation of each forcing. The discussion on advantages and limitations of the hydrological-hydrodynamic framework is addressed in ***Comment 3***. Suggested correction:

> "**Forcings and boundary conditions.** The hourly VCSN rainfall (mm/h) is spatially masked using the rainfall mask polygon and imposed as rain-on-grid in the model. The river injections are prescribed as hourly timeseries of discharge hydrographs (m^3/s) and imposed as areal source terms ($(m^3/s)/m^2$) uniformly distributed over a 10 m×10 m square surface centred on the corresponding injection point with no horizontal velocity. Inland domain boundaries are treated as impermeable walls to avoid water exiting the domain upstream of the injection points. The 500 m buffered region around the lower catchment is added to the simulation domain to ensure that the hydrological and rainfall forcings occur well inside the domain to prevent numerical instabilities due to water accumulations at the inland solid wall boundaries…"

4.1 Baseline modelling and validation

Comment 8: *Section 4.1 states that flood extents are qualitatively assessed by comparison with post-event aerial and satellite imagery. Since these data could potentially be used to derive binary inundation maps, please explain why no quantitative flood-extent performance metric, such as the Critical Success Index (CSI), was used, or discuss any limitations that prevented such an analysis.*

Response: Direct binary inundation maps could not be systematically derived from the available imagery using traditional remote sensing methods due to the lack of imagery records during the event. In the Heretaunga Plains, binary maps derived from post-event aerial imagery and in-situ reports were shared by the Hawke's Bay Regional Council. Despite being informative, the maps were produced as a scoping tool rather than a precise account of the event, and they were found unsuitable for direct model validation as they sometimes misrepresented flood extents with errors exceeding hundreds of metres with unidentified inland runoff. The production of suitable digitized binary maps at regional scale would have demanded work and expertise exceeding the capabilities of this project. We therefore decided to adopt a "visual"/qualitative approach at the cost of not providing quantitative model performance metrics such as CSI index for the flood extent prediction.

6.1.4 Grid refinement map

Comment 9: *The statement that representing the upper reaches at 16 m resolution has generally small effects (typically lower than 0.5 m water depth) and negligible influence on downstream floodplain hydrodynamics appears rather strong. Please clarify the basis for this conclusion. Was it derived from a grid-sensitivity analysis, previous validation studies, or operational experience? In addition, some information on the typical dimensions of the river channels represented at 16 m resolution would help readers assess the suitability of this grid spacing.*

Response: The use of 16 m resolution in the upper reaches was based primarily on operational modelling experience and inspection of the resulting flow patterns, rather than on a formal grid-convergence analysis. In the Hawke's Bay and Tairāwhiti-Gisborne domains, the upper reaches channels are typically 10–30 m wide and are therefore only partially resolved at 16 m resolution (one to two cells across). However, they are generally carved within steep, well-defined gullies such that the surrounding topography remains sufficiently well represented to constrain flows to the correct downstream pathways. Previous internal modelling projects have also shown limited sensitivity of downstream floodplain hydrodynamics to resolutions up to 32 m where the channel geometry and major terrain discontinuities remain adequately represented. We agree that the original statement was too strong and will revise the manuscript to clarify that the small influence is an empirical modelling observation rather than a formally quantified uncertainty range. A systematic sensitivity analysis combining 8 m, 16 m and 32 m resolutions with plausible variations in surface roughness would provide a more rigorous assessment of resolution uncertainty, however this extends beyond the scope of this manuscript. Published systematic flood modelling datasets at national and continental scale typically operate at 30 m to 100 m resolution (Wing et al., 2024; Dottori et al., 2022; Wing et al., 2017; Sampson et al., 2015). To our knowledge, the finest resolution published for a systematic regional or national scale flood dataset of comparable spatial coverage is uniform 2 m for Austria (Blöschl et al., 2024; Buttinger-Kreuzhuber et al., 2022). The 4–32 m adaptive grid used here is consistent or finer than most systematic datasets and offers an appropriate trade-off between resolution and computational constraints. Suggested correction in the discussion:

> "**The Grid refinement maps** were designed to balance accuracy and computational cost based primarily on operational modelling experience and inspection of the resulting flow patterns, rather than on a formal grid-convergence analysis. Levees are resolved at 4 m resolution, providing better representation of crest elevation and overtopping. Along the major rivers without levees, 8 m resolution was considered sufficient to capture channel shape and floodplain runoff sources. Upper channel reaches can be represented at 16 m and occasionally at 32 m resolution. These channels are typically 10–30 m wide and are therefore only partially resolved at 16 m resolution. However, they are generally carved within steep, well-defined gullies such that the surrounding topography remains sufficiently well represented to capture the hydraulically important topographic features and constrain flows to the correct downstream pathways. Previous internal modelling projects have also shown limited sensitivity of downstream floodplain hydrodynamics to resolutions up to 32 m where the channel geometry and major terrain discontinuities remain adequately represented. Downstream narrow channels and urban drainage (width ~10–20 m) can also be represented at 16 m resolution. At this resolution, topography smoothing can favour early spilling. However, most of these channels were already overtopped at the peak of the event, and terrain smoothing contributed only locally to total inundation. Preliminary sensitivity tests also indicated that differences generally remained within ±0.5 m water depth and had limited influence on flood map representation in severely impacted areas, however in-depth assessment is required to formally estimate impacts. In farmlands, where drainage and irrigation channels are dense, this uncertainty is compounded by unopened culverts and unreferenced waterway obstructions, making their respective contributions difficult to isolate. Overall, uncertainties associated with the refinement map are considered moderate with moderate impacts in the regions of interest."

6.2 Technical validation

Comment 10: *As stated by the authors in the Introduction, more than 30 levee breaches occurred along the two major rivers of the Heretaunga Plains during Cyclone Gabrielle. These failures were likely to play a major role in controlling flood propagation across the protected floodplain. The manuscript states that breach locations and timings were provided by the Hawke's Bay Regional Council and that the topography was dynamically adapted during the simulation to reproduce the observed failures. However, this important aspect of the modelling framework is only briefly described, yet it appears to be one of the key components controlling the flood dynamics. Please provide more details on how the dynamic modification of the DEM was implemented. In particular, it would be useful to clarify how the breach geometry was represented (e.g. crest lowering and final breach width), how the temporal evolution of the breaches was incorporated into the simulations, and whether the adopted representation was based on field observations or on simplifying assumptions. It would also be helpful to provide some information on the typical height of the levees above the surrounding floodplain, to better appreciate the hydraulic significance of the simulated breaches.*

Response: A detailed paragraph dedicated to the description of the levee implementation will be added to the models calibration description. Suggested addition.

> "**Levee breaching.** In the Heretaunga Plains, calibration aimed to integrate the levee breaches along the Ngaruroro River and Tūtaekurī River, two major rivers of the water catchment. Reasonable information about the locations and timing of the breaches were collated and shared by HBRC, as summarised in Fig. 5. The breaches were here parameterized using simple pre-defined time-varying geometry deformations (Dewals et al., 2011; Vacondio et al., 2016) to make use of the available information. Each breach was defined by a deformation start time, deformation duration and gridded geometry deformation map representing the total elevation variation over the deformation duration. The deformation is triggered once the simulation time reaches the deformation start time. The ground elevation is then linearly modified in each cell until reaching the final elevation variation prescribed in the deformation map. Breaching start times were set consistently with Fig. 5. Even though no precise record of breaching duration was available, information shared by HBRC

reported sudden levee bursts and rapid widening of the breaches, with final deformation typically occurring within the hour. This observation was extended to all breaches, and the breaching duration was set to a unique 30 min period for all breaches. Locations and final shape of the breaches were manually digitised from the post-event aerial imagery and used as mask to create the deformation map. For each breach mask, the deformation mask was defined to achieve a smooth surface across the levee breach width."

Comment 11: *The Introduction states that numerous bridges were washed away during the event. However, I could not find a description of how bridges were represented in the hydrodynamic model. Please clarify whether bridge structures (and their possible obstruction effects) were explicitly considered in the simulations or neglected. A brief discussion of the implications of this modelling choice would help readers better understand the limitations of the proposed framework.*

Response: Waterway infrastructures captured by the remote sensing such as bridges were fully removed during the DEM hydraulics conditioning, it is mentioned in section "3.2. Topography reconstruction". No additional conditioning was systematically applied to the DEM or surface roughness to account for residual impacts due to the presence of these infrastructures. The description of the hydraulic conditioning in Sect. 3.2 will be updated to clarify the treatment. A brief discussion of the implications on the output will be included in the limitations and internal variabilities section under "DEM condition". Suggested additions:

Sect. 3.2: "The raw DEMs were then hydraulically conditioned to open waterways and remove obstructions captured by the remote sensing elevation data (e.g. thick vegetation covering waterways). Hydraulic infrastructures were also fully opened (e.g. bridges, culverts) and were considered to have no direct impact on the flow dynamics. This is justified by the large variability in geometrical features of the infrastructures and the generally high local uncertainty introduced by the hydraulic conditioning."

Discussion: "The full opening of all hydraulic infrastructures is a rather strong assumption. Under critical design flow conditions, smaller bridges and culverts are likely to introduce non-residual pressure loss on channel flows. Ignoring their impact, whether through the surface roughness or hydraulic connector models, likely acts towards enhancing channel transmissibility and reduce overtopping. Effectively, the flow conditions during the event far-exceeded critical design conditions and the geometric design of the opening in the DEM is considered as the associated dominant source of uncertainty, residual drag being second order in comparison."

Comment 12: *Table 3 contains useful information on the different simulation scenarios and their associated confidence levels. However, it is neither adequately introduced nor discussed in the text. Consider adding a brief description of the purpose of the table and highlighting its main contents to help readers interpret the information provided. In particular, could the authors clarify the meaning of the "Glass wall scenario" and the statement "Levees are made impermeable"? Would terms such as non-erodible or non-overtoppable better describe the intended modelling assumption? In addition, Table 3 indicates that, at least for the Waiapu catchment, the calibrated model was obtained by modelling the entire catchment with "no river injections". This appears to differ substantially from the general modelling framework described in Section 3. Could the authors clarify whether this configuration may effectively correspond to a catchment-scale rain-on-grid hydrodynamic simulation? If so, it would be useful to briefly discuss the rationale behind this alternative modelling strategy and its implications.*

Response:

1. A dedicated section will be added in the discussion to introduce the confidence ratings and summarise the dataset quality and calibration approach. Suggested addition:

"Consistent with the uncertainty classification in Table 3 and the simulation validation, each scenario modelled is assigned a confidence level classified as low/moderate/high according to the level of uncertainty and performance of the models. Generally, the overall satisfactory representation of the flooding characteristics by the baseline models in the Hawke's Bay catchments (moderate to high confidence) indicates that the framework provides reliable output given accurate enough rainfall input and no local flow conditions influenced the hydrodynamics in sections of the catchment (e.g. breaches, large sediment load). Simultaneous modelling of a large terrain extending across two regional districts also enabled the identification of characteristic limitations of the VCSN rainfall estimate to consistently represent severe storms such as Cyclone Gabrielle over large spatial distances due to uneven density of rainfall gauges. In line with the previous points, calibration of the hydrological forcings would ideally aim to correct the rainfall input itself. Such work would be the topic of an entire parent study requiring the use of additional quantitative (e.g. additional citizen gauges) and qualitative (e.g. residents accounts, expert interpretation and judgement…) input with sufficient granularity to satisfactorily correct local rainfall patterns. Hydrodynamic simulations produced using this modelling framework can then be used to verify if the reconstructed rainfall produces representative flood hazards.

The calibration strategy for hydrological forcings was kept simple and scalable to balance project timeline and cover most relevant catchments. Yet, it led to meaningful improvements in the representation of flood extents and peak flow conditions, which are key quantities for design practice, exposure, risk and damage analysis. In catchments where hydrological forcings were mostly due to rainfall in the upper catchments, tuning of the injections resulted in meaningful improvement of the model performances at minimal computational costs. Catchment-scale rainfall calibration and modelling in the Waiapu catchment led to acceptable model performance given the scarce validation data. This came however with large hydrodynamic computation overhead due to the inflated grid cell count (~50% computation runtime using one Tesla P100 GPU). Catchment scale calibration was acceptable as exploratory practice for a single catchment. However, this approach was impractical as a systematic calibration strategy, as the cumulative computational cost would have exceeded the project timeline. This constraint is rapidly diminishing with continued GPU hardware and computational methods progress. Two alternative calibration strategies were also considered but not explored. (i) Joint tuning of both rain-on-grid and river hydrographs in the simulation domain showed to be complex and convoluted by introducing an additional layer of discussion to justify the relative tuning weight placed on the rain-on-grid and river injections. (ii) Tuning of rainfall over the entire water catchment and creation of new sets of injection hydrographs through re-running the hydrological model exceeded the objectives of the calibration which targeted solely the hydrodynamic model inputs."

2. The "glass wall scenario" is indeed designed to prevent overtopping and assess flow characteristics (water depth, velocity) if river flows were confined within the main waterways. The description is updated accordingly. Suggested changes:

"Glass wall scenarios**.** The crest elevation of all levees identified in the NZIS database was increased by 5 m to prevent overtopping and confine the flow within waterways."

3. The calibrated model for the Waiapu catchment was indeed constructed as a catchment-scale rain-on-grid. Description of the rationale behind this design choice will be added to the Calibration section. Discussion of the implication of catchment-scale vs hybrid hydrological-hydrodynamic modelling framework is addressed in the Discussion section in line with ***Comment*** *7*. Suggested change of the Waiapu calibration description:

Calibration description: "In Waiapu, the hydrodynamic domain covers a large portion of the total water catchment. Part of the rainfall peaks are included in the simulation domains (see Fig. 1*a*) and the rain-on-grid contributed to a non-negligible portion of the total flow at the lowest river reaches. Initial tests showed that scaling of the injection hydrographs alone did not result in consistent improvement of the flood extent throughout the domain. Instead of the joint calibration of injections and rain-on-grid, a second catchment-scale model covering the entire water catchment was created (no injections) and input rainfall was calibrated over this area. The VCSN hyetographs at all grid cells were incrementally adjusted using a uniform factor α_r until satisfactory predictions were obtained. While increasing the computation time by ~50% (25h up to roughly 40h), this approach offered a simple, consistent and scalable processing framework applicable under time constraints for a single catchment."

Technical corrections

Response: We thank the referee for flagging these editorial corrections. Technical corrections will be directly addressed at manuscript resubmission and the ones requiring explanations are detailed below.

Figures

• Please add the sub-panel labels referenced in the captions (a, b, …) to Figures 1 and 4, as these are currently not shown in the figures.

Response: Corrected. The sub-panels labels have been added

• *Please clarify in the Figure 2 caption that the white areas in Figure 2b correspond to regions outside the Hawke's Bay and Tairāwhiti regional districts.*

Response: Corrected. The caption has been updated accordingly.

• *Please check the spelling of "Tairāwhiti" in the Figure 2 caption, as it appears inconsistently formatted compared to the rest of the manuscript.*

Response: Corrected. Tairāwhiti is now used consistently throughout the manuscript.

• *In Figure 4b, could the green legend entry correspond to the 500 m buffer zone rather than the simulation domain? The current labelling appears ambiguous and may deserve clarification.*

Response: The figure will be updated to clarify the meaning of the polygons

• *Figure 7. Consider presenting the regional flood map with north oriented upwards. Although this may require a slightly different layout, it would improve readability and facilitate comparison with the other maps presented in the manuscript.*

Response: Corrected. The map has been rotated

• *Figure 8b. The predicted inundation map contains numerous small, isolated flooded cells distributed across the domain. Please clarify whether these features represent natural topographic depressions, shallow ponding generated by the rainfall forcing, or other modelling artefacts. If they correspond to very small water depths, consider applying a minimum water-depth threshold in the visualization to improve the readability of the inundation map.*

Response: These isolated flood cells are known numerical artefacts of the current hydrostatic reconstruction implemented in BG-Flood. They can be filtered during post-processing, but we have retained them in the published outputs for transparency. Clear mention will be made in the document.

• *Figure 9c. The text refers to the "black circles" identifying the levee breaches at Dartmoor and Fernhill. However, these locations are difficult to identify in the figure. Please consider making these markers more evident (e.g. larger symbols or arrows) and/or adding the corresponding location labels to improve readability for readers unfamiliar with the study area.*

Response: Location of the breaches is now introduced in a separate figure, and clear markers identify the locations mentioned in the text.

Text

• *The acronym VCSN appearing in the Figure 1 caption is not defined at first occurrence. Please define it explicitly (e.g. Virtual Climate Station Network) in the text and/or in the figure caption for readers unfamiliar with New Zealand climate datasets.*

Response: Corrected. VCSN is now defined upon first encounter

• *Line 179. The sentence describing the Brookfield bridge piers appears incomplete ("due to the large accumulation of ..."). Please revise.*

Response: Corrected.

• *Line 193. Consider replacing "The hydrographs (m^3/s)" with "The discharge hydrographs", as the units are implicit in the variable being described and the revised wording is more precise.*

Response: Corrected

• *Lines 289–291. Please ensure consistency in the notation used throughout the manuscript. The water surface elevation is previously denoted as ηs and ηs,max, whereas in this section it is referred to as zs and zsmax.*

Response: η_s has been replaced by z_s consistent with the dataset variable name.

• *Line 307. Please correct the figure reference: "Fig. 6" should read "Fig. 5".*

Response: Corrected.

• *Line 363. Please correct the typo "uncertinity" to "uncertainty".*

Response: Corrected.

• *Line 412. Please check the figure reference. The text refers to Fig. 9a and 9c, whereas the corresponding KDE plots are shown in Fig. 10a and 10c.*

Response: Corrected.

• *Summary and conclusions. This section includes not only a summary of the main findings and conclusions, but also an extensive discussion of possible extensions and future developments of the modelling framework. Consider adopting a title such as "Summary, conclusions and future work" (or similar), which would better reflect its content.*

Response: With the reorganisation of the manuscript and the inclusion of a dedicated discussion section, the conclusion is now a proper summary and conclusions.

References

Blöschl, G., Buttinger-Kreuzhuber, A., Cornel, D., Eisl, J., Hofer, M., Hollaus, M., Horváth, Z., Komma, J., Konev, A., Parajka, J., Pfeifer, N., Reithofer, A., Salinas, J., Valent, P., Výleta, R., Waser, J., Wimmer, M. H., and Stiefelmeyer, H.: Hyper-resolution flood hazard mapping at the national scale, Nat. Hazards Earth Syst. Sci., 24, 2071-2091, 10.5194/nhess-24-2071-2024, 2024.

Buttinger-Kreuzhuber, A., Waser, J., Cornel, D., Horváth, Z., Konev, A., Wimmer, M. H., Komma, J., and Blöschl, G.: Locally Relevant High-Resolution Hydrodynamic Modeling of River Floods at the Regional Scale, Water Resour. Res., 58, e2021WR030820, 10.1029/2021WR030820, 2022.

Crawford-Flett, K., Blake, D. M., Pascoal, E., Wilson, M., and Wotherspoon, L.: A standardised inventory for New Zealand's stopbank (levee) network and its application for natural hazard exposure assessments, J. Flood Risk Manag., 15, e12777, 10.1111/jfr3.12777, 2022.

Dewals, B., Erpicum, S., Detrembleur, S., Archambeau, P., and Pirotton, M.: Failure of dams arranged in series or in complex, Natural Hazards, 56, 917-939, 10.1007/s11069-010-9600-z, 2011.

Dottori, F., Alfieri, L., Bianchi, A., Skoien, J., and Salamon, P.: A new dataset of river flood hazard maps for Europe and the Mediterranean Basin, Earth Syst. Sci. Data, 14, 1549-1569, 10.5194/essd-14-1549-2022, 2022.

Revell, M., Carey-Smith, T., and Moore, S.: A severe frontal rain event over the lower North Island of New Zealand with intense embedded convective banding, Weather and Climate, 39, 14-27, 10.2307/26892909, 2019.

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., and Freer, J. E.: A high-resolution global flood hazard model, Water Resour. Res., 51, 7358-7381, 10.1002/2015WR016954, 2015.

Smart, G. M.: Improving flood hazard prediction models, Int. J. River Basin Manag., 16, 449-456, 10.1080/15715124.2017.1411923, 2018.

Tait, A., Henderson, R., Turner, R., and Zheng, X.: Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface, Int. J. Climatol., 26, 2097-2115, 10.1002/joc.1350, 2006.

Vacondio, R., Aureli, F., Ferrari, A., Mignosa, P., and Dal Palù, A.: Simulation of the January 2014 flood on the Secchia River using a fast and high-resolution 2D parallel shallow-water numerical scheme, Natural Hazards, 80, 103-125, 10.1007/s11069-015-1959-4, 2016.

Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., and Erickson, T. A.: Validation of a 30 m resolution flood hazard model of the conterminous United States, Water Resour. Res., 53, 7968-7986, 10.1002/2017WR020917, 2017.

Wing, O. E. J., Bates, P. D., Quinn, N. D., Savage, J. T. S., Uhe, P. F., Cooper, A., Collings, T. P., Addor, N., Lord, N. S., Hatchard, S., Hoch, J. M., Bates, J., Probyn, I., Himsworth, S., Rodríguez González, J., Brine, M. P., Wilkinson, H., Sampson, C. C., Smith, A. M., Neal, J. C., and Haigh, I. D.: A 30 m Global Flood Inundation Model for Any Climate Scenario, Water Resour. Res., 60, e2023WR036460, 10.1029/2023WR036460, 2024.