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

1. General comments:

Comment 1: *This manuscript presents a dataset of hydrodynamic flood inundation simulations reconstructing the flood hazard characteristics associated with ex-Tropical Cyclone Gabrielle across 16 locations in Aotearoa New Zealand. The dataset is potentially valuable due to its high spatial resolution, regional coverage, and integration of hydrological and hydrodynamic modelling workflows. Validation against hydrographs, aerial imagery, and surveyed high-water marks further strengthens the potential usefulness of the dataset.*

Response: We are glad that the referee recognises the usefulness of the dataset.

Comment 2: *While the dataset itself is potentially valuable, the manuscript does not sufficiently articulate its specific scientific and practical relevance beyond the reconstruction of a past flood event. Section 1 focuses extensively on the description of the flood event and associated impacts, but the broader applicability, reuse potential, and added value of the dataset relative to existing flood inundation products remain insufficiently discussed. In particular, Sections 1 and 8 would benefit from a clearer and more concrete discussion of the intended applications, target users, and unique contributions of the dataset.*

Response: We will add a dedicated section discussing the general relevance and applicability of the dataset in the regional and national context of Aotearoa New Zealand as well as current applications.

Comment 3.1: *Additionally, the manuscript would benefit from substantial editorial revision. Several Sections (3,4,6) are difficult to follow due to verbose writing, excessive subsection fragmentation, and the mixing of conceptual workflow descriptions with implementation details. Reducing the number of subsections and improving the conceptual organization of the methodology would substantially improve readability and reproducibility. For example, the distinction between Sections 4 and 6.2 is not always clear, as validation methodology, calibration procedures, validation results, and uncertainty interpretation are distributed across multiple sections. This fragmentation makes the overall modelling and evaluation workflow difficult to follow. A clearer separation between methodology, validation protocol, and validation results/discussion would substantially improve readability.*

Response: We agree with the referee's comment. The manuscript is reorganised to tighten the presentation and enhance information delivery. Unnecessary subsection headings are also removed. The proposed new structure is

1. Background summary
2. Study area

3. Methodology
    3.1. Domain creation
    3.2. Topography reconstruction
    3.3. Gridded rainfall
    3.4. Hydrology
    3.5. Hydrodynamics
    3.6. Validation strategy
    3.7. Model calibration and simulated scenarios
4. Dataset records
5. Technical validation (previously 6.2)
6. Discussion
    6.1. Uncertainty sources, model limitations and scenarios confidence (previously 6.1)
    6.2. Dataset quality: Table 3 + discussion
    6.3. Dataset usage and framework applications
7. Data availability
8. Summary and conclusions

Comment 3.2: *Finally, some aspects of the calibration procedure require stronger methodological justification, particularly regarding the scaling of rainfall and river inputs and the dynamic modification of topography to introduce levee breaches, as these choices may substantially influence the simulated inundation patterns and validation results.*

Response: The description of the calibration procedure will be updated to include a clear presentation of the reasons motivating the different calibration approaches (rainfall and river injection scaling, levee breaching treatment). Suggested additions:

> "A second modelling phase is undertaken to address discrepancies in the baseline results. The purpose of the calibration was to reconstruct flow conditions and flood extents with sufficient confidence to inform population and asset exposure, damage/cost assessment, potential design, decision making and scenario analysis. Given the limited information available, the uncertainties in many of the inputs (e.g. rainfall, geomorphology, sediment and debris transport) and the regional coverage of the models, calibration was carried out as fine-tuning of the hydrodynamic models to correct for both known effects and significant uncertainties where sufficient validation data was available to ensure reliable corrections through two pathways: correction of the hydrological forcings and inclusion of levee breaching.
>
> Validation data was sparse: six domains contain no river gauge stations, only two count more than five stations in their lowest sections, and surveyed high-water marks are available in three domains already well covered with gauges. Consequently, detailed calibration is carried out only for six domains where sufficient data ensured reliable corrections.
>
> **Hydrological forcing tuning.** In the Northern domains where hydrological forcings were identified as the dominant sources of error (see rainfall uncertainties discussion), calibration focused on tuning the river injection hydrographs and rainfall inputs
>
> In Wairoa and Tolaga Bay, the hydrodynamics in the lower catchments were mostly driven by the river flow forcings (rainfall in the upper catchment) and calibration focused on tuning the river injection hydrographs. Due to the large number of injections (> 100 in each domain), a scalable and easily applicable tuning strategy was considered. The injection hydrographs were scaled by a uniform or spatially targeted factor α_{inj} which was incrementally adjusted through a series of simulations until satisfactory representation of the flood extent and gauge hydrographs predictions were obtained. However, this came with the trade-off in accurately capturing the rising and receding limbs of the gauge hydrographs.
>
> - In Gisborne, calibration efforts aimed at correcting major model deficiencies observed along the Te Arai River, tributary of the Waipaoa River, in the South-Eastern part of the domain. Scaling of the river hydrographs feeding into the Te Arai River led to a satisfactory representation of the Te Arai

River overtopping its banks and maximum water depths at the surrounding surveyed high-water marks.

- 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. 1a) 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 did not result in consistent improvement of the flood extent throughout the domain and calibration of the rainfall was necessary. We chose to calibrate rainfall over the entire catchment using a catchment-scale rain-on-grid model of the Waiapu water catchment. 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 and scalable framework computationally applicable for a single catchment.

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.

Local calibration of the Esk Valley. The Esk Valley, located in the Aropaoanui domain, was among the most severely impacted during the event and the most challenging to calibrate. Calibration efforts focussed on a subdomain covering the lower Esk Valley (Fig. 8b) and aimed to adjust the boundary hydrograph at the valley inlet to compensate for the inherent deficiencies of the numerical solver. Owing to the small size of the computation domain, the simulation was run on a uniform 8 m grid.

In some cases, additional models are derived from the best-performing models (baseline or calibrated) to investigate location specific processes (e.g. impact of inundation protection wall along the Tukituki River). The list of scenarios stored in the datasets are summarized in Table 2."

2. Specific comments:

***Comment 4 - P2L49:** The term "spectacular flash floods" appears informal and subjective for scientific writing. A more precise and neutral description of the event severity would be preferable.*

Response: The term "spectacular flash floods" is replaced with "extreme flooding".

***Comment 5 - P2L52-53:** The statement regarding "multi-scale impacts" is currently too broad and would benefit from greater specificity. The manuscript should clarify what scales and impact types are being referred to and how infrastructure failures demonstrated these effects.*

Response: The section was reformulated to make explicit the progression of damage from local to regional scale. Suggested correction:

"Failure of critical infrastructure emphasized the vulnerability of the regions to flood hazards and revealed impacts cascading up to regional scale. Levees breaches at more than 30 locations along the two major rivers of the Heretaunga Plains locally intensified flooding in already severely impacted areas (Hawke's Bay Regional Council et al., 2023). At regional scale, multiple roads and bridges were washed away and, in Tairāwhiti-Gisborne, 90% of the communication towers were out of service during the peak of the event, further isolating the population and blocking emergency response routes. In Hawke's Bay, 50% of telemetered rainfall and river stations lost power, and many river gauges attached to bridges that collapsed were lost (Hawke's Bay Regional Council et al., 2023), resulting in temporally and spatially fragmented records of the event for operational forecasting and post-event recovery."

Comment 6 - P3: *Several acronyms used in Figure 1 (e.g., VCSN, HBRC, and GDC) are not defined upon first use, which may reduce readability for an international audience unfamiliar with regional New Zealand institutions and datasets. These abbreviations should be expanded either in the main text or in the figure caption.*

Response: Manuscript is updated to systematically introduce acronyms upon first encounter.

Comment 7 - P3: *In Figure 1 (a) and (b) is missing.*

Response: Corrected.

Comment 8 - P6L106: *The rationale and methodology used to select the modelling domains are currently insufficiently described. In particular, the use of "population" and "economical exposure" as selection criteria requires clearer definition. Additionally, the term 'economical exposure' appears inappropriate in this context and should likely be replaced.*

Response: This was just an opening sentence with the greater proportion of the explanation coming later. But we do acknowledge that it is unclear. We are now moving this in with the domain generation section so that all the information is provided together. We suggest replacing this paragraph with the following which will then lead into the explanation.

> "Sixteen domains were identified that cover the main floodplains of the two regions. These were first manually identified by drawing an 'area of interest' polygon around the floodplain. These were then refined into a lower catchment and associated wider catchment using the method outlined below. The domains are listed in Table 1, and the wider catchments are presented in Figure 1b. The land extent of lower catchments varies from 6 km2 to 820 km2, with an average area of 275 km2, the associated wider catchment areas vary from 27 km2 to 6025 km2, with an average of 1650 km2."

Comment 9 - P7: *The title "Numerical framework" appears somewhat misleading, as the section describes the overall modelling workflow and data-processing pipeline rather than the numerical framework itself.*

Response: The title "Numerical framework" is replaced with "Modelling workflow"

Comment 10 - P7L111: *Several acronyms and institution names (e.g., NIWA) are used without being defined upon first use. This reduces readability for an international audience and should be addressed consistently throughout the manuscript.*

Response: See response to **Comment 6**. The manuscript is updated to systematically introduce acronyms upon first encounter.

Comment 11 - P7L120: *The opening description of the domain delineation workflow in Section 3.3 is difficult to follow. In particular, the phrase "through the compilation of geographical data" remains too vague, and the specific role of the different datasets (e.g., shoreline, LiDAR coverage...) in refining the area of interest polygon is not clearly explained.*

Response: We will reword this for clarity and ensure that we explain the role of all the datasets. Suggested rewording:

> "Each domain is created by refining an area of interest polygon using the method described below. This method incorporated river catchment data extracted from the River Environment Classification v2.5 database (REC2.5) (Earth Sciences New Zealand, 2022), LiDAR coverage, population and building density and shoreline data. REC2.5 is a digital river network which compiles the river reaches, their associated sub-catchments and

> other collated geographic and topographic information. For each domain, four polygon masks are created (Fig. 4): the water catchment, the lower catchment, the rainfall mask and the hydrodynamics simulation domain."

> "...aggregating sub-catchments as identified in REC2.5 within the area of interest polygon that have LiDAR coverage and also with bed slope < 20°, population densities > 0.00004 inhabitant/km2 and building densities > 0.00008 building/km2. The domain is extended to the shoreline on the seaward side. The upstream borders of the lower catchment ..."

> "The hydrodynamics simulation domain is created by applying a 500 m buffer around the rainfall mask. The buffer ensures that all rainfall and the river injection points are well inside the simulation domain which avoids numerical stability issues caused by forcing too close to the boundary (see Sect. 3.7.2)."

Comment 12 - P9L155: *A few lines above, the framework is introduced as "GeoFabrics", while here it is written as "Geofabrics". Consistent naming should be used throughout the manuscript.*

Response: The manuscript will be double checked for consistency and updated accordingly.

Comment 13 - P9L162-165: *This sentence is difficult to follow and would benefit from being split into shorter statements.*

Response: The sentence was reformulated and split into two clear statements for clarity. Suggested rewording:

> "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."

Comment 14 - P11L226: *Repetition (see line210).*

Response: The paragraph was updated accordingly. The presentation of the advantages of the mesh's block-uniform quadtree structure is concentrated in the introductory paragraph. This specific paragraph now only focuses on the mesh generation procedure.

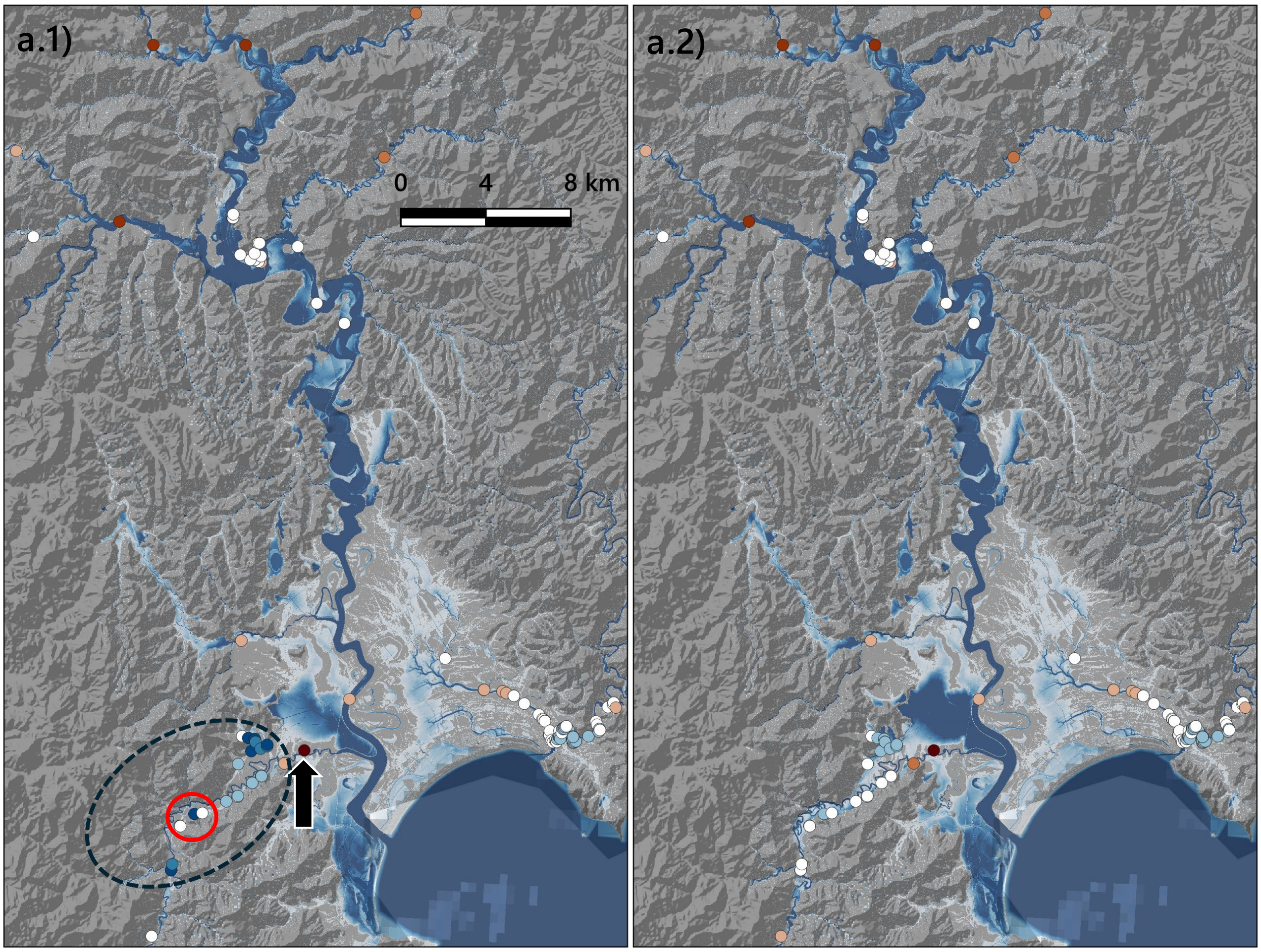
Comment 15 - P12L236: *Acronym already introduced.*

Response: See **Comment 6** - Corrected

Comment 16 - P13L261: *The treatment of surveyed points that are dry in the model outputs is potentially problematic. Replacing dry cells with the nearest wet grid cell may artificially reduce prediction errors and obscure flood extent mismatches. This procedure requires stronger methodological justification and its potential influence on the reported validation statistics should be explicitly assessed.*

Response: At survey points located in predicted dry grid cells, the model error is evaluated by comparison with the water surface elevation at the nearest wet grid cell. This provides a meaningful metric of the vertical error at the boundary of predicted inundated zones: it quantifies how far the predicted flood surface falls from the survey point elevation and captures model underprediction in a way that treating dry cells as an "immediate fail" would not. Two main justifications support the procedure. Firstly, at 4–8m resolution, the nearest wet cell is typically within one to three grid cells – a spatial offset of 4–20m – which is well within the positional uncertainty of the HWM survey itself. Less than 10 survey points in the Gisborne model are located more than 30 m from the nearest wet grid cell. Secondly, in areas with rapidly varying terrain, a predicted dry cell at a survey point does not necessarily represent model failure in the accepted sense, but may simply reflect terrain smoothing caused by the discretisation, especially where the scale of local topographic gradients exceeds the grid resolution.

A large flood extent mismatch remains generally identifiable with this approach, and manifests as a meaningful negative error, as seen along the Te Arai River for the Gisborne base model (dashed circle in the Figure below). As the referee correctly notes, the method may prove ineffective and artificially reduce prediction errors in certain pathological configuration such as survey points near topographic depressions or shallow surface runoff that create isolated wet cells disconnected from the main inundated area. Failure of the approach was found at only 5 points along the Te Arai River for the Gisborne base model (e.g. red circle in the Figure below) which is considered satisfactory. The alternative, treating all dry cells as an 'immediate fail', would artificially inflate the error count. This rationale will be integrated into the manuscript validation methods.



Comment 17 - P12L238: *The title 'Modelling strategy' appears somewhat misleading, as the section primarily describes calibration, validation, and scenario simulations rather than the modelling strategy itself.*

Response: The title has been modified alongside the general structure of the manuscript. This section now sits in the methods section.

Comment 18 - P13: *The calibration methodology requires further clarification and justification. While the manuscript describes the calibration as 'minimal', the applied adjustments include scaling rainfall inputs, modifying river injections, and dynamically altering topography to introduce levee breaches, which may substantially influence the simulated inundation patterns. The criteria used to apply these corrections, their magnitude, and their potential impact on model reproducibility and event-specific tuning are currently insufficiently documented.*

Response: See **Comment 3**.2.

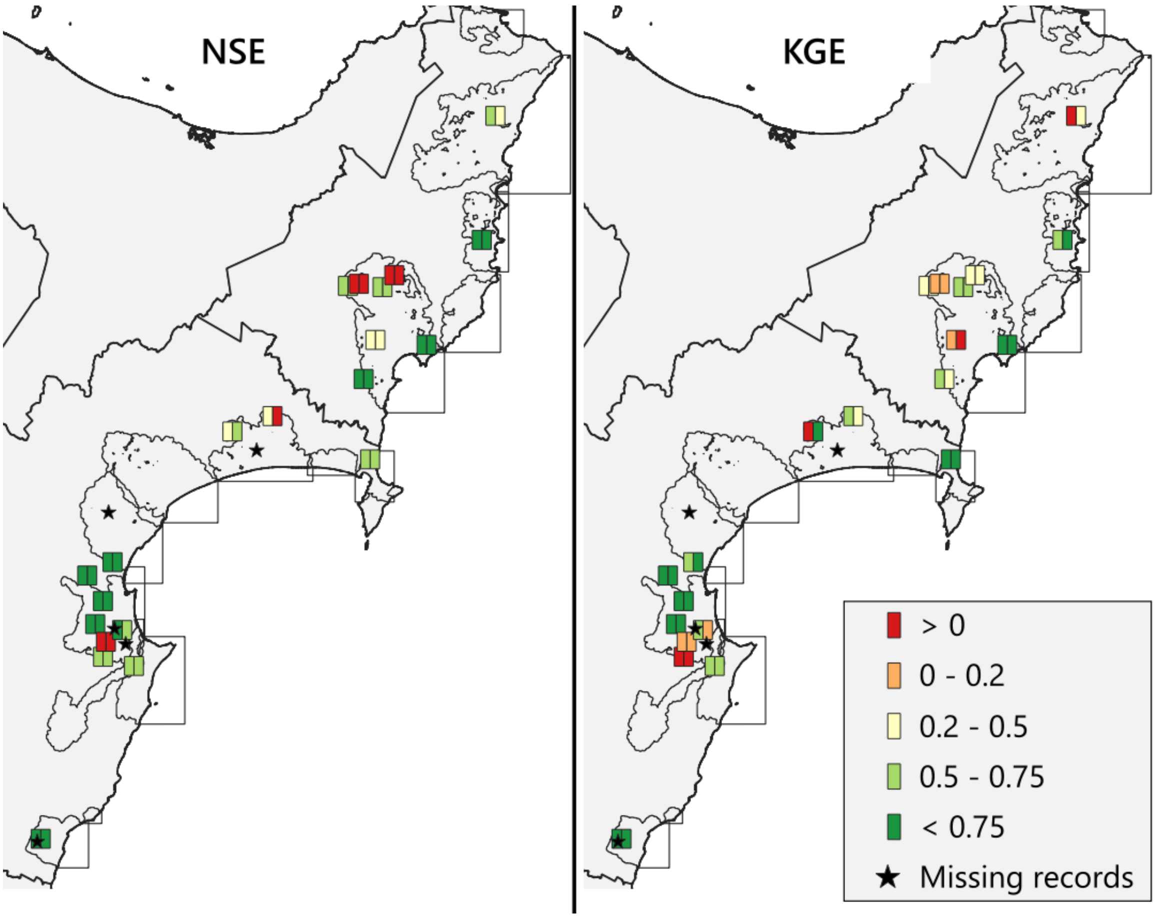
Comment 19 - P20: *Figure 8d is very hard to read. The manuscript states in the abstract that the models were "validated against hydrograph records"; however, the corresponding validation procedure remains insufficiently documented in the main text. While example comparisons are presented in this figure, it is unclear which quantitative criteria or metrics are used to assess model performance.*

While the high-water mark validation procedure is described in detail, the corresponding validation approaches based on hydrograph records and aerial imagery remain comparatively under-documented. Although example comparisons are presented, it is often unclear how these validations were performed in practice, which variables or criteria were used, and whether the assessment was qualitative or quantitative.

Response:

1. Fig. 8 is now split into 2 separate figures to enhance readability.
2. 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, they were found unsuitable for direct model validation as they sometimes misrepresented flood extents with errors exceeding hundreds of metres. 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 for the flood extent prediction.
3. We initially compared the water depth hydrographs with the recorded water surface elevation. This was a difficult task as the exact coordinates of the gauges and vertical datum at the gauges were rarely available and difficult to deduce with confidence. Consequently, we now present a comparison of the discharges at gauges where the information is available using two characteristic model skill metrics: the Kling-Gupta Efficiency (KGE) (Gupta et al., 2009) and Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970). The metrics are computed at 21 stations where sufficient records between the 2023-02-12 12:00:00 and 2023-02-15 12:00:00, and the results are summarised in the table below. The skill metrics clearly highlight the locations where model deficiencies were already identified, mostly in Tairāwhiti. In the Heretaunga Plains, the low metrics values correspond to narrow channels that were submerged in surface runoff and cannot be considered representative of the models' performances. A more targeted discussion will be integrated to the manuscript. Figure 7 will also be updated accordingly to present discharges rather than water levels.

	Base model		Calibrated			Base model		Calibrated	
NSE	No. of stations	[%]	No. of stations	[%]	KGE	No. of stations	[%]	No. of stations	[%]
> 0.75	9	43 %	8	38 %	> 0.75	6	29 %	9	43 %
0.5 – 0.75	6	29 %	7	33 %	0.5 – 0.75	7	33 %	2	10 %
0.2 – 0.5	3	14 %	2	10 %	0.2 – 0.5	2	10 %	5	24 %
0 – 0.2	0	0 %	0	0 %	0 – 0.2	3	14 %	3	14 %
< 0	3	14 %	4	19 %	< 0	3	14 %	2	10 %



Comparison of the models skill at the river gauges measured using Nash-Sutcliffe Efficiency (NSE) and Kling-Gupta Efficiency (KGE) for the base models (left half of the boxes) and, where applicable, the calibrated models (right half of the boxes). In domains where no calibration was performed, the right half is also the base model.

References

Dewals, B., Erpicum, S., Detrembleur, S., Archambeau, P., & Pirotton, M. (2011). Failure of dams arranged in series or in complex. *Natural Hazards*, *56*(3), 917-939. https://doi.org/10.1007/s11069-010-9600-z

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1), 80-91. https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.08.003

Hawke's Bay Regional Council, Wairoa District Council, Hastings District Council, Napier City Council, Central Hawke's Bay District Council, & Toitū Te Whenua – Land Information New Zealand Data Service (LINZ). (2023). *LiDAR dataset, Hawke's Bay, New Zealand 2020-2021* (https://doi.org/10.5069/G9S75DH2

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282-290. https://doi.org/https://doi.org/10.1016/0022-1694(70)90255-6

Vacondio, R., Aureli, F., Ferrari, A., Mignosa, P., & Dal Palù, A. (2016). 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*(1), 103-125. https://doi.org/10.1007/s11069-015-1959-4