I was invited to review the manuscript by Mohamed et al. (2025), which presents an integrated deep learning (DL) framework trained on outputs from a physics-based morphodynamic model to efficiently accelerate spatiotemporal predictions of river–floodplain evolution. While the framework does not aim to introduce new physical concepts or process understanding, it effectively leverages deep learning to reproduce the behavior of the reference physics-based model in a computationally efficient manner. Consequently, the primary contribution of the manuscript lies in methodological advancement and workflow efficiency rather than in the development of new morphodynamic theory. Overall, the study is interesting and promising. Nevertheless, several points would benefit from clarification or further refinement. Addressing these comments to the satisfaction of the editor and reviewers would, in my view, strengthen the manuscript, and I would be happy to recommend it for publication.

1. The main focus of this study is to demonstrate that deep learning (DL), trained on outputs from the physics-based model, can reproduce comparable results at a substantially higher computational speed. It does not explore the broader potential of DL for simulating hydrodynamic and morphodynamic processes in river or floodplain systems. <u>Therefore, the main contribution lies in computational efficiency rather than providing new physical insights</u>. In addition, numerous studies have applied deep learning to achieve higher computational speeds; for example, Synthetic in Bentivoglio et al. (2022, *Deep learning methods for flood mapping: a review of existing applications and future research directions*) and Karim et al. (2023, *A review of hydrodynamic and machine learning approaches for flood inundation modeling*). Consequently, the statement in the abstract (line 34) claiming that this study represents "pioneering new frontiers in fluvial morphodynamic modelling" is misleading and should be revised.

2. The abstract reports good performance metrics (e.g., RMSE, R^2), which is encouraging. However, it should explicitly clarify that the DL model is trained on outputs from the physics-based model rather than observed data, as its performance reflects replication of the physics-based model rather than independent validation against real-world measurements.

3. Regarding the 2D geomorphological simulation using the physics-based HEC-RAS model, which serves as the reference for the DL model, the physical representation appears relatively simplified, and many key parameters are missing. For example, the Manning's n values, the sediment transport formulations employed (e.g., bedload: van Rijn, 1984; Engelund and Fredsøe, 1976; Meyer-Peter and Müller, 1948) are not specified. It is also unclear how sediment input at the upstream boundary (bedload and/or suspended load) is specified, and how the time steps are configured during the simulation. Furthermore, calibration and validation of the physics-based model appear to be missing. Without these essential steps, it is difficult to assess the reliability of the HEC-RAS model for simulating morphodynamic processes in the study area, which raises concerns about the robustness of the DL model that relies on it as reference data.

4. Similarly, in lines 113–115, the authors state: "*The resulting framework predicts the dynamics of essential hydrodynamic outputs: water depth and flow velocity, which are fundamental inputs to the morphodynamic target, represented in bed change maps*." and the bed change results appear to rely solely on these two hydrodynamic outputs. This is insufficient for morphodynamic modelling. In addition to hydrodynamic variables, morphodynamic processes also depend on sediment-related

variables, such as sediment concentration (suspended load, which plays an important role in sedimentation within the floodplain), total sediment transport and transport rates (bedload and/or suspended load), and sediment properties (e.g., grain size and settling velocity).

However, I acknowledge that addressing these points (3 and 4) would require considerable time, access to observed data, and expert knowledge in morphodynamic modelling, particularly with a thorough understanding of the study area. Given that the main aim of this study is to demonstrate the ability of the DL model to achieve substantially higher computational efficiency, I leave it to the editor to decide whether these concerns need to be addressed for publication.

5. Data training (lines 144–145): Approximately 11 flood events were used for training, 3 for validation, and 6 for testing, with this information presented only visually in Figure 1b. This approach is rather simple, and it is unclear how this selection was justified. It would therefore be helpful if the authors could include a sentence explaining the rationale for this choice—for example, whether the 11 training events cover the full range of flood magnitudes (from small to extreme events) in terms of probability, total water volume, or peak discharge over the period of record, or at least whether the validation and testing events fall within the range of the training data (quantified). Providing such quantitative information would help readers better understand the representativeness of the training dataset

6. Line 297-300. In physics-based models, error accumulation is a major concern, as small errors at early stages can grow and propagate over time. Here, the CB-trained framework demonstrates rapid recovery after the first testing event and shows no apparent accumulation of relative errors in subsequent events (Fig.6), the reasons for the initially high relative errors are not discussed. Further clarification on the mechanisms underlying this recovery would help readers better understand the robustness and stability of the framework in continuous simulations.

7. Section 3.4: The authors test a range of time steps for the DL model, from 0.25 hr to 16 hr. However, a key detail is missing: the time step used in the physics-based reference model is not provided. This information is important because the physics-based model serves as the baseline for training and evaluating the DL model. For example, if the physics-based model uses a 30-minute or 5-hour time step, it could strongly influence which DL model time step is most appropriate. Once the time step of the physics-based model is provided, it would be useful to discuss the optimal DL time step in relation to the physics-based model, as well as the trade-offs between computational efficiency and prediction accuracy.

8. Lines 420–421 mention Manning's n, grain-size distributions, and cohesive properties, which as come out of nowhere (see point 3).

9. Given the complexity of geomorphological processes, more discussion on the limitations and broader applications of DL is needed, especially since it is purely simulation-based. The DL model is evaluated only against the physics-based HEC-RAS model, not observed field data, meaning any errors in HEC-RAS are inherited. It should be noted that the DL model reflects its ability to replicate HEC-RAS rather than providing independent validation of real-world morphodynamics. The study focuses on a relatively simple domain, a single reach with one boundary condition. It would be valuable for the authors to discuss applying DL models to more complex river systems, such as large-scale river–floodplain networks with multiple boundaries. Additionally, the discussion could address the model's handling of internal morphological changes (e.g., dredging, dam construction, or localized sediment

management) and the lack of explicit incorporation of expert knowledge or physical constraints. Outlining strategies to address these limitations would guide applications of DL models to more realistic, actively managed river systems

Minor comments: The use of the Morphological Acceleration Factor (Mf) requires careful consideration as it can affect flood-wave propagation and sediment dynamics. In this study, for example, the duration of discharge exceeding 600 m^3s^{-1} for event Tr-1 is ~10 hours when Mf = 1, but reduces to ~2 hours when Mf = 5 (Figure 1b). With Mf = 1, high discharge persists long enough to propagate downstream and spread onto the floodplain. In contrast, Mf = 5 shortens the high-flow duration, potentially reducing downstream peak discharge due to increased floodplain storage and attenuation. Since sediment transport is strongly non-linear with discharge, these changes in flood-wave dynamics may introduce systematic errors in sediment transport and bed-change simulations. <u>The authors should also highlight the importance of expert knowledge in morphological modelling, not only for physics-based models but also for DL models.</u>