Journal of Hydrology 665 (2026) 134578

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

The mirage of the silver bullet: Exploring the limitations of high-resolution data in flood model validation

Pratik Chakraborty *, Christophe Dessers, Pierre Archambeau, Michel Pirotton, Sébastien Erpicum, Benjamin Dewals

Hydraulics in Environmental and Civil Engineering (HECE), University of Liège (ULiège), Quartier Polytech 1, Allée de la Découverte 9, Liège, 4000, Belgium

ARTICLE INFO

This manuscript was handled by Emmanouil Anagnostou, Editor-in-Chief, with the assistance of Giulia Sofia, Associate Editor.

Keywords:
Flood model validation
Hydrodynamics
Validation
Uncertainty
Post-flood survey
Critical success index
High-water marks

ABSTRACT

Validation of inundation models presents persistent challenges, particularly in urban floodplains where structural complexity often exacerbates discrepancies between model outputs and observation datasets. Findings from the present study suggest that while availability of high-resolution field data can enhance validation efforts, it is not the final piece of the puzzle. The study highlights the (often inseparable) complex combination of observed data limitations, model uncertainties, and structural discrepancies between model and observed datasets, which strongly influence validation outcomes. Using the July 2021 Vesdre Valley flood in Belgium as an illustrative framework, the research evaluates the performance of a high-resolution 2D hydrodynamic model (WOLF). The unprecedented detail of the post-flood survey provides a unique opportunity for rigorous validation. Four so-called 'reconciliation methods' are explored to address structural discrepancies (in post-processing) between observation data and computational outcomes, highlighting the effect of the choice of method. At the highest model resolution (2 m), the critical success index (= 0.86) indicates strong spatial agreement across sectors, with a mean absolute error of 0.56 m in modelled maximum flood depths. By proposing and critically evaluating various methods for reconciling inherent differences between observational and computed datasets, this study highlights the complexity of model validation beyond data availability. Additionally, it offers recommendations for refining post-flood survey methodologies to minimise uncertainties associated with the validation process.

1. Introduction

Climate change is intensifying extreme weather events, including floods, which pose growing threats to lives, infrastructure, and economies (Merz et al., 2021; Blöschl et al., 2019; Alfieri et al., 2017; Hirabayashi et al., 2013). Recent catastrophic floods in Europe, such as the DANA floods in Valencia, the July 2021 mega-floods in Germany, Belgium and the Netherlands, highlight the severe impacts of these phenomena (Rodríguez Castro et al., 2025; Bertola et al., 2023; Ludwig et al., 2023; Tradowsky et al., 2023). As efforts towards flood mitigation increase, hydrodynamic modelling, particularly 2D numerical modelling, has become central to flood-risk assessment studies, providing valuable information for scenario analysis, and development of mitigation measures and adaptation strategies.

While very-large-scale hydrodynamic models (e.g., at national or continental scales) have a growing presence in literature, their grid resolution constraints inhibit their utility for high-fidelity flood hazard estimation (Shen and Jiang, 2023). High-resolution models (grid cell size < 10 m) at local scales are more feasible. But with increasing access to high-performance computing (HPC) facilities and emergence of GPU-based solvers (Khosh Bin Ghomash et al., 2024), even basin-scale modelling at such high-resolutions appear computationally feasible and practicable. However, Yalcin (2020) highlights a saturation of gains via their findings - that hyper-resolution models (grid cell size < 2 m) are not necessary to obtain reliable outcomes. The author discusses the comparable performance of models with 2 m and 10 m computational grid cell sizes. However, in the study the sustained level of performance at lower resolutions is attributable to the sub-grid modelling capabilities of the HEC-RAS modelling tool that was used (Casulli and Stelling, 2011), which accounts for the details of an underlying higher resolution DEM. For other models without such sub-grid modelling, the resolution of the DEM and the computational grid are the same. Accordingly, Alipour et al. (2022) highlight that when requiring accurate estimates of flood depths, DEM/model resolution is the single most critical factor.

Urban flood modelling or large-scale models that contain urbanised areas present additional challenges. The treatment of buildings in flood

* Corresponding author.
E-mail address: p.chakraborty@uliege.be (P. Chakraborty).

https://doi.org/10.1016/j.jhydrol.2025.134578
Received 29 July 2025; Received in revised form 10 October 2025; Accepted 6 November 2025
Available online 20 November 2025

inundation modelling has been discussed at length in literature. Different approaches exist for incorporating buildings into hydrodynamic models, each with its own assumptions and limitations. Schubert and Sanders (2012) highlight four different building treatment methods (BTM). One approach is the *building-resistance* method, which simplifies the representation of buildings by assigning a high resistance parameter to cells within building footprints. This method is easy to implement and computationally efficient but does not capture the detailed interaction of flow with building geometries. Alternatively, the *building-block* method represents buildings as solid blocks by raising the ground elevation to the height of rooftops. This approach accounts for the basic geometric obstruction of buildings on flood flow, but can introduce under-prediction in flood extents if the mesh resolution is not sufficiently fine to capture flow paths between buildings. The *building-hole* method attempts to represent the detailed shape of buildings by creating holes in the computational mesh aligned with building walls, with free-slip wall boundary conditions. This method can be computationally expensive and requires pre-processing of building shapes, particularly in complex urban environments. Finally, the *building-porosity* method offers a balance between accuracy and computational efficiency by introducing spatially distributed parameters like porosity and a building drag coefficient to model the impact of buildings on flooding without explicitly resolving their exact geometries. Each of these methods introduces a different type of uncertainty into the modelling, with choices impacting model accuracy, computational cost, and data requirements.

Rigorous and robust validation of hydrodynamic models remains rare, both in practice and in scientific literature. This is especially true for validation with respect to floodplain inundation. A variety of data sources can contribute to validation, each addressing different facets of model performance. Water-level gauges provide high-frequency temporal records at fixed points, making them well suited for evaluating hydrograph timing and peak discharge. But they offer limited spatial coverage, often restricted to being instream and close to hydraulic structures (Werner, 2004). Remotely sensed flood extents, such as those derived from optical or radar imagery, allow large-scale assessment of inundation boundaries, yet they lack information on water depth and are often restricted by cloud cover, revisit times, and water–vegetation interactions. Work from the past decade (Scarpino et al., 2018) indicate the potential of Synthetic Aperture Radar (SAR) to provide extent validation data as well as spatially distributed water depth data but are known to acquire data with significant noise (Bates, 2023). Crowd-sourced observations and photographic evidence can provide valuable context on event dynamics, but they are typically sparse and heterogeneous and therefore often are a complementary data category in practice.

Against this backdrop, High-Water Marks (HWMs) remain uniquely important as they capture spatially distributed information on maximum flood depths and extents. Their utility lies in the ability to directly link water surface elevation to ground elevation at specific locations, thereby providing a critical benchmark for validating simulated inundation depths.

However, HWMs are not without significant constraints. Their positional accuracy is influenced by uncertainties in both elevation and location: small errors in surveying or in associating the water line with the ground level can propagate into discrepancies in derived flood depths. In addition, HWMs are more frequently recorded in accessible, built-up environments, such as on building facades, than in open fields or vegetated areas. This uneven distribution can skew the dataset towards certain land-cover types or urban settings, thereby limiting its representativeness of catchment-wide flood behaviour. Finally, HWMs provide only a snapshot of the peak water level. They cannot indicate the sequence of flooding, the duration of inundation, or transient dynamics such as short-lived overtopping or backflow. As a result, if validating the temporal dynamics of a flood model is the goal, HWMs fall short as a data source.

Nevertheless, Table 1 presents several studies that validate flood models against observed field data. Mignot et al. (2006) validated their hydrodynamic model against 99 HWMs for the 1988 flood event in Nîmes, France. The domain extents being fairly limited (1.4 km^2) allowed for street level refinement of the mesh. The HWMs reported were along building facades - along which the computational cell edges were aligned. However, since the model computed water depths at cell centres only, the comparison was carried out using the corresponding cell centres in the domain. Smith et al. (2021) used a combination of 131 HWMs for a 2008 flood event in Sugar Creek basin in the USA to valid their model. The authors explore a so-called areal-sector approach for maximum flood depth comparison owing to the uncertainty linked with measurements (as stated by Shustikova et al. (2019) as well) and the accuracy of close-field mesh. Wing et al. (2021), on the other hand, carried out simulations at a very-large-scale, studying 9 different events. The outcomes of validation for a 2019 event in Nebraska, USA, which recorded 1023 HWMs, are presented in Table 1. de Arruda Gomes et al. (2021) alluded to the uncertainties related to post-flood interventions in a basin which have an impact on the base data of simulations, while Kastridis et al. (2020) highlight that failure of flood-defence mechanisms not accounted for in the models can have an effect on model outcomes.

The scarcity of robust validation is often attributed to the limited availability of field data (Bates, 2023; de Moel et al., 2015; Alarcón et al., 2024; Jacob et al., 2020), as large-scale data collection during or after flood events is particularly resource-intensive (Alarcón et al., 2024; Molinari et al., 2019). When data are collected, their utility depends significantly on their density, spatial coverage, and type (Bhola et al., 2019). Surveys yielding dense datasets are typically localised (Segura-Beltrán et al., 2016; de Moel et al., 2015), whereas those covering larger areas often feature sparse data points and/or sources with higher levels of uncertainty. For instance, basin-scale post-flood surveys frequently collect only flood extent data. While water-level measurements (e.g., HWMs) may also be recorded, their spatial density is often low due to the logistical and resource demands of such efforts.

This recurring narrative within flood-modelling literature positions data-rich post-flood surveys as the ‘silver bullet’ — the key to overcoming the final hurdle before reliable model validation can become the norm. But, while the data-rich nature of post-flood surveys can be of immense value, there are several other facets of the model validation process that ought to be taken into account.

For instance, several studies, including very recent findings of Khosh Bin Ghomash et al. (2025) (who deal with the extreme flooding in the Ahr valley of Germany in 2021), state as their general recommendation the importance of representing buildings as physical obstacles in hydrodynamic models, as in the building-block and building hole BTMs. Given that the latter suffers from the inability to allow for building overtopping as well as difficulties related to meshing, the building-block BTM emerges as the most suitable choice for hydrodynamic modelling of extreme flooding especially when the domain is inclusive of complex urban environments. Now, when using this approach, buildings and other tall structures appear as elevated impermeable topography in the numerical model. This implies that the area corresponding to the footprint of a building is not inundated. This is not the case for inundation maps reconstructed using observation points since they are mostly based on interpolation, taking only bare-earth terrain into account. This leads to differences between modelled and observationally reconstructed flood maps that hinder direct comparability. The vast majority of flood model validation studies exhibit a notable lack of clarity regarding the procedures employed (or lack thereof) to bridge these differences prior to the computation of validation metrics.

To that end, this study proposes four different post-processing methods of reconciling these differences between flood maps derived from numerical models and field observations (when using the building-block BTM) before the computation of validation metrics. These methods are applied and tested for a hydrodynamic modelling exercise at

Table 1
Instances of validation studies based on flood extents and/or maximum flood depths.

Reference	Hydrological domain	Hydrodynamic domain	Mesh Resolution	Obs. Count	CSI Score[a]	Depth error[b] (m)	Remarks
Mignot et al. (2006)	13.5 km^2	1.4 km^2	25 m	99	–	0.41	Hightlight the effect of street network resolution on water depth predictions.
Smith et al. (2021)	679 km^2	–	10 m	131	–	0.58	Report and analyse uncertainties associated with HWM-based evaluation.
Wing et al. (2021)	–	2,500 km^2	30 m	1023	0.80	1.12	Refinement in regions with available fine-resolution terrain data.
de Arruda Gomes et al. (2021)	7,260 km^2	29 km^2	10-30 m	–	0.66	–	Report uncertainties due to topography data version.
Kastridis et al. (2020)	31.36 km^2	0.26 km^2	5 m	46	0.82	0.35[c]	Report underestimation of flooding in residential areas due to flood defence failure in actuality.
Shustikova et al. (2019)	–	50 km^2	25 m	46	0.81	0.61[c]	Report accuracy range of ±0.5 m for HWMs.
Segura-Beltrán et al. (2016)	118 km^2	5 km^2	1 m	64	0.76	0.64[c]	Report the localised nature of HWM data with respect to total flooded area.
Chakraborty et al.[d]	**685 km^2**	**12 km^2**	**2 m**	**12,293**	**0.86**	**0.56**	Unprecedented post-flood survey after a mega-flood.

[a] Best-case value reported when multiple simulations are run.
[b] Mean absolute error (MAE) unless indicated otherwise.
[c] Root mean squared (RMSE) error.
[d] Present study.

the scale of a valley (~50 km). The selected domain comprises densely built urban environments and open floodplains devoid of structures, providing a comprehensive testing ground for the analysis. Further, the availability of a post-flood survey with unprecedented detail (including more than 12,000 HWMs) provides a unique framework for validation. At the same time, the study also reports the effects of uncertainties linked to mesh resolution, digital elevation model (DEM) resampling, and their implications for flood extent and maximum flood depth validation.

2. Data and methods

2.1. Case study

The floods of July 2021 in north-west Europe were among the most devastating natural disasters to hit the region in decades. The flood was triggered by the Bernd low-pressure system, which caused extreme and intense rainfall in parts of Germany, Belgium and the Netherlands (Journée et al., 2023). In Belgium, the Vesdre catchment was among the worst affected (Fig. 1). The valley experienced large-scale inundation of floodplains all along the Vesdre river as well as its tributaries, such as the Hoëgne. Densely populated areas such as Verviers and Pepinster experienced severe damage to infrastructure as well as loss of life (Rodríguez Castro et al., 2025; Dewals et al., 2021). Flow-rates, water levels and their corresponding gradients also far-exceeded records (Chakraborty et al., 2024; Archambeau et al., 2022b,a). Several tens of bridges are also know to have been clogged by debris, leading to varying degrees of structural damage and/or influence on local flood inundation and flow dynamics (Erpicum et al., 2024).

In the aftermath of the flood, mapping flood extents posed significant challenges, especially in urban areas. While aerial imagery provided an initial understanding, it was insufficient for detailed delineation of inundation in complex urban environments like Verviers and Pepinster (Zischg et al., 2018). To address this, the regional public service authority, Service public de Wallonie (SPW), along with other public offices, mobilised 180 agents and carried out a large-scale field survey in the month following the mega-flood. This effort collected over 30,000 data points. Evidences of flood extents and water depths were documented across 39 affected municipalities (Anon, 2021). The ground survey was initiated 3 weeks after the recession of the flood-waters and is known to have continued for another 3 weeks. However, the exact distribution of collection of HWMs within this timeline is unknown due to a lack of timestamps in the publicly available version of the dataset. The HWMs were predominantly recorded from the facades of buildings and other infrastructure. For instance, in the case of the municipality of Verviers, where the highest number of data-points are found in the domain of the present study, roughly 98% of all data-points correspond to buildings and other infrastructure and only 2% come from observations made on undeveloped land. When recording observations from buildings, it is also known that the ground survey personnel only took measurements from the exterior facades of the buildings. The geolocation of an observation, however, was done only to roughly correspond to a building itself. Therefore, the points of geolocation are frequently located within the buildings themselves while it is known that no measurements internal to the buildings were made.

From the 30,000 data points, the SPW retained 16,000 points after quality checks - 12,293 of which are within the domain of the present study itself (Fig. 2). The SPW further proceeded to prepare a continuous raster product conveying flood extents and depths (Figure S1). This was created using an **inverse distance weighted (IDW)** interpolation method combined with a high-resolution digital terrain model (DTM) (v2013-14, vertical accuracy ±0.12 m). The SPW reports the IDW interpolation to have had a power parameter of 3 and a fixed search radius of 450 m. In some reaches with missing observation points, the nearest observation point(s) were used to carry out linear extrapolation. Notably, the parameterisation and the final outcome were products of a combination of data sources including arial imagery from multiple helicopter flights, the ground survey, remote sensing products such as those from the L'Institut royal météorologique (IRM), TerraFlood, Rapid Mapping Copernicus etc. In the absence of access to the various other sources of information stated above, the authors consider it both appropriate and necessary to utilise the official public product of the SPW, since its independent re-creation or validation is not feasible and also lies outside of the scope of the present study.

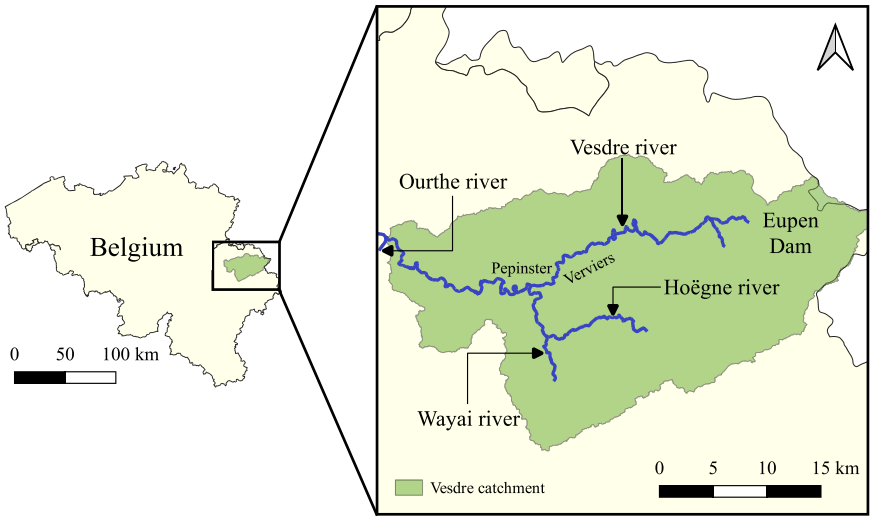


Fig. 1. The Vesdre catchment in Belgium.

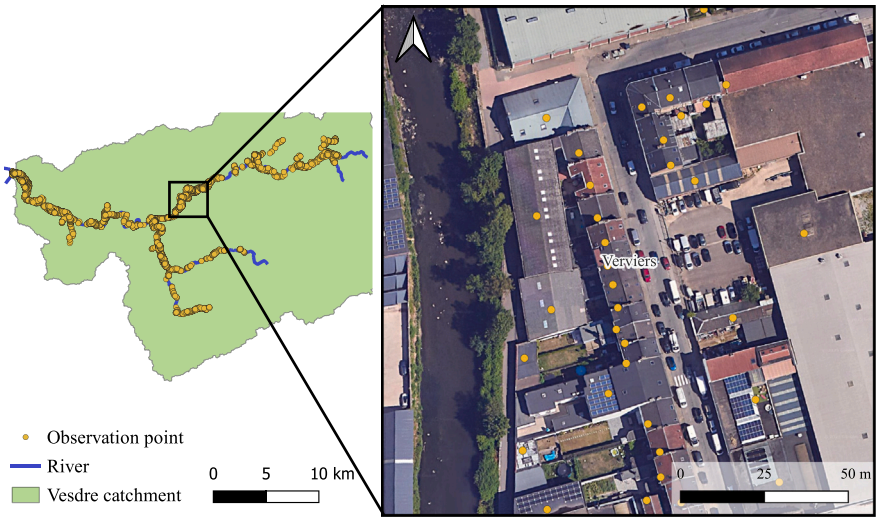


Fig. 2. Distribution of field observations along the Vesdre, Hoegne and Wayai rivers, and a zoom on the Verviers municipality.

Representing a significant event with ample data for rigorous investigation, the Vesdre valley flood of 2021, in Belgium, serves as the focal case in the present study.

2.2. Numerical modelling

2.2.1. Model domain

In the present study, the entire the Vesdre river, as well as the Hoegne and Wayai tributaries, are explicitly modelled. Contributions from other smaller tributaries are accounted for using internal injection points on the main river. Finally, this results in a significantly large simulation domain - starting from the Eupen dam and culminating at the confluence of the Vesdre with the Ourthe river approximately 50 km downstream.

2.2.2. Numerical modelling setup

A 2D depth-averaged hydrodynamic model is used for hindcasting of the Vesdre Valley flood event. The GPU-based hydrodynamic solver employs the fully dynamic two-dimensional shallow water equations. The inflows in the hydrodynamic model were estimated from rainfall data using a physically based, spatially distributed runoff model. Rainfall from the RADFLOOD21 (version 27/06/2024, IRM) dataset was used to force the hydrological model. The hydrological model was calibrated against available reservoir water level time-series and hydrographs available at gauging stations. Further details related to the hydrological modelling may be found in Dessers et al. (2025). The integrated modelling system is developed at the University of Liège (https://wolf.hece.uliege.be/).

For stability reasons, the hydrodynamic time step was constrained by the Courant–Friedrichs–Levy (CFL) condition, with the CFL number set to a value of 0.4. Additionally, the model's handling of wetting and drying of computational cells has been detailed in Supplement S2. While this is a challenge, many robust algorithms have been developed for handling wetting and drying of computational cells in the resolution of the shallow-water equations (for e.g., (Begnudelli and Sanders, 2007; Casulli, 2009; Liang and Marche, 2009)). The algorithm used herein involves solving the continuity equation by iteratively adjusting the outflow fluxes from the computational cells in which drying occurs – which is similar to the 'flux minimisation method' described by Rulot et al. (2012) for modelling sediment transport over a partially rigid bottom. The model was extensively verified through theoretical test cases and validated against laboratory data from experimental tests involving wetting and drying of the topography (Erpicum et al., 2010). The model was also used in multiple studies involving inundation modelling with complex, real-world topographic configurations (Mustafa et al., 2018; Ernst et al., 2010; Beckers et al., 2013).

Table 2
Simulation codes based on grid resolution and resampling operator.

	2 m	5 m	10 m
Minimum	2m-min	5m-min	10m-min
Mean	2m-mean	5m-mean	10m-mean
Maximum	2m-max	5m-max	10m-max

Computational meshes (uniform orthogonal cells) at three different resolutions have been tested: 2 m × 2 m, 5 m × 5 m, and 10 m × 10 m, comprising approximately 3 million active cells at the highest resolution. High-resolution topographic and bathymetric data (LiDAR v2021-22; 0.5 m × 0.5 m) were resampled using three distinct operators: minimum, mean, and maximum. It may be clarified that in the rest of the manuscript any mention of the term DEM refers to a modified Digital Surface Model (DSM) product specifically curated for hydrodynamic simulations. This includes the retention of infrastructure such as buildings and the removal of obstructions like bridge decks, trees etc., which are also reflected in a typical DSM. Bare-earth terrain models are, on the other hand, consistently referred to as DTMs. Nonetheless, this combination of resampling operators and resolutions resulted in a total of nine simulations (Table 2), enabling a simultaneous sensitivity analysis of the model to grid resolution and resampling methods. It ought to be stated that the testing of the minimum and maximum operators is not intended as an investigation of the general consensus pertaining to the use of a mean-based operator. Instead, they are intended as bounding cases to demonstrate the degree to which systematic exaggeration or suppression of local gradients can affect flood model predictions.

Boundary conditions at the downstream limit were specified using a Froude number-based approach to ensure numerical stability and consistency with flow dynamics. Spatial variations in surface roughness and flow resistance were incorporated through variable Manning's coefficients derived from land-use data (Figure S2). Finally, the building-block BTM is adopted.

2.3. Model validation

2.3.1. Flood maps reconciliation method

As discussed previously, one of the key challenges in the comparison of the observed inundation raster with modelled one, when using the building blocks BTM, lies in the treatment of built-up areas. A hypothetical scenario is presented to elucidate the point. In Figs. 3a and 3b, the raw modelled and observed inundation maps are shown. For the sake of explanation, it is assumed that the outer margins of the two inundation extents are in good agreement. The modelled inundation raster (Fig. 3a) shows three isolated non-inundated patches. While one of them is due to a naturally elevated topography, the others are due to the presence of buildings. On the other hand, the observed inundation raster only shows the region of elevated topography as not inundated while being continuous through buildings. Naturally, to compare the two maps one should consider the footprint of such buildings in the modelled map to be inundated as well.

An obvious requisite is therefore a clear definition of what is implied by 'building inundation'. In this work, a building is regarded as inundated if water can plausibly propagate into its footprint from any surrounding area. Since the propagation of the flow through the buildings is not always explicitly represented, this may be operationalised through a practical proxy wherein a building footprint is considered inundated whenever at least one of its neighbouring cells is inundated. The proxy definition carries the underlying assumption that water from the surrounding inundated cell(s) could propagate into the building footprint in the absence of the impermeable building-block. The proxy definition is clearly conservative in that regard, since any surrounding cell being inundated may not imply internal propagation (which would be contingent on the horizontal and vertical positioning of openings such as doors, windows etc.).

Accordingly, four different methods for reconciling observed and modelled inundation maps are introduced and compared:

In the **buffered-building (BB) method**, any building polygon that has a water depth greater than a user-defined threshold in a surrounding cell is considered to be inundated.

In principle, a 'surrounding cell' should be the ones immediately next to the building polygon. However, this is found to not always be the case. This is because the grid on which the DEM is registered cannot always capture the true shape of a building (Muthusamy et al., 2021; Savage et al., 2016). This can be due to non-alignment of the grid and the sharp outline of a building (as depicted in Fig. A.1). As a result, the elevation difference between the top of a building and its immediate surroundings can get smoothened in a DEM. This implies that in such cases the DEM bears a larger footprint of the buildings than in reality. Depending on the resolution of the DEM or the resampling operation used, this smoothening can be substantial or limited.

If so, then the modelled inundation waterline along the outline of the deformed building can be 'some' distance away from the actual building outline. Hence, it is possible that upon checking for water depths in the immediate bordering cells of a building polygon (actual building outline) one finds no inundation and falsely marks it as 'not inundated'. In order to deal with these concerns, a buffer layer is introduced to the 'true' building polygon (Fig. 3c). For a sufficient buffer size, these enlarged (building) polygons may now be expected to intersect with the inundation waterline when formerly they were not.

A second, more simplified, **filling-of-holes (FoH) method** may also be considered. In this method, 'holes' in the model inundation map are filled indiscriminately. A 'hole' in the inundation map (Fig. 3e) is defined by a region in the raster, comprising one or several cells, that are marked as non-inundated (inundation status → 0) whilst being completely surrounded by inundated cells (inundation status → 1). To 'fill' these holes, therefore, means to switch the inundation status of all the cells in such patches to '1'. In Fig. 3e, it is also understood as to what would not be considered as a hole and therefore would not be filled. In the same figure the indiscriminate nature of the FoH method is highlighted, where the elevated and non-inundated region is filled since it technically qualifies as a hole. This contradicts the objective of only compensating for holes in the computed inundation map that are due to the building-block BTM.

In the two aforementioned approaches the idea has been to process the modelled inundation raster so as to ensure its comparability with the observed raster.

In contrast, the **inverse adaptation (IA) method** looks to modify the observed values raster for comparison with the modelled one. This is done by removing inundated cells from the observed raster which are underneath building polygons (Fig. 3h). Notably, this is a much simpler operation; all '1' (or '0') values that are within a building area are turned to (or kept as) '0'. This is consistent with the idea that in the building-blocks BTM, one does not expect inundation in cells where there are buildings. This approach helps in bypassing some of the struggles of the aforementioned methods, viz.:

- Buildings no longer have to be identified as being inundated or otherwise. In the BB approach, it was necessary to do so because the footprints of inundated buildings were being 'added' to the total inundation area.
- Holes due to high topography are maintained in the observed inundation raster, since the processing never fills/adds values to the observed raster (which was a shortcoming of the FoH approach).

However, the IA method may lead to underestimations if a flood of sufficient magnitude inundates low-elevation buildings and auxiliary structures, including garages and single-story constructions.

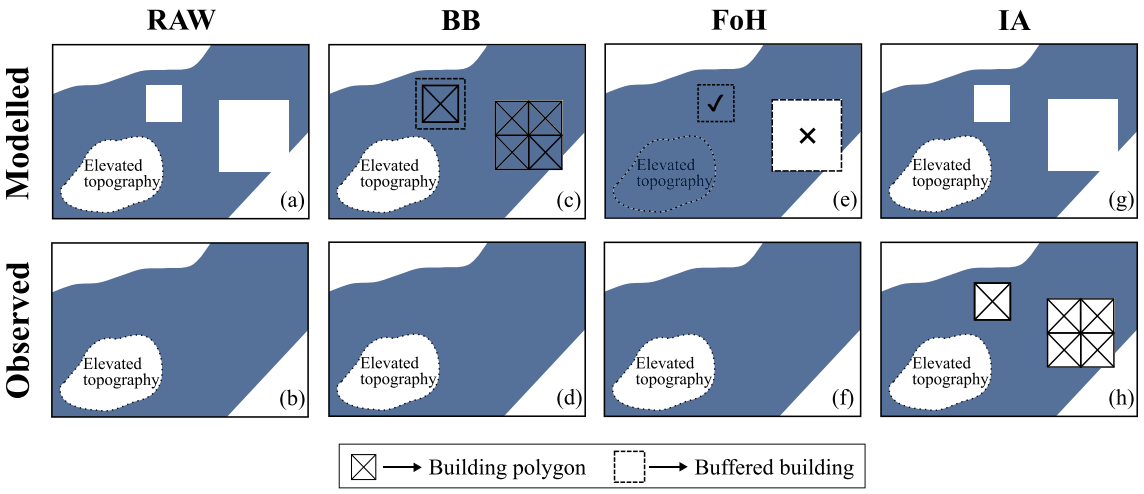
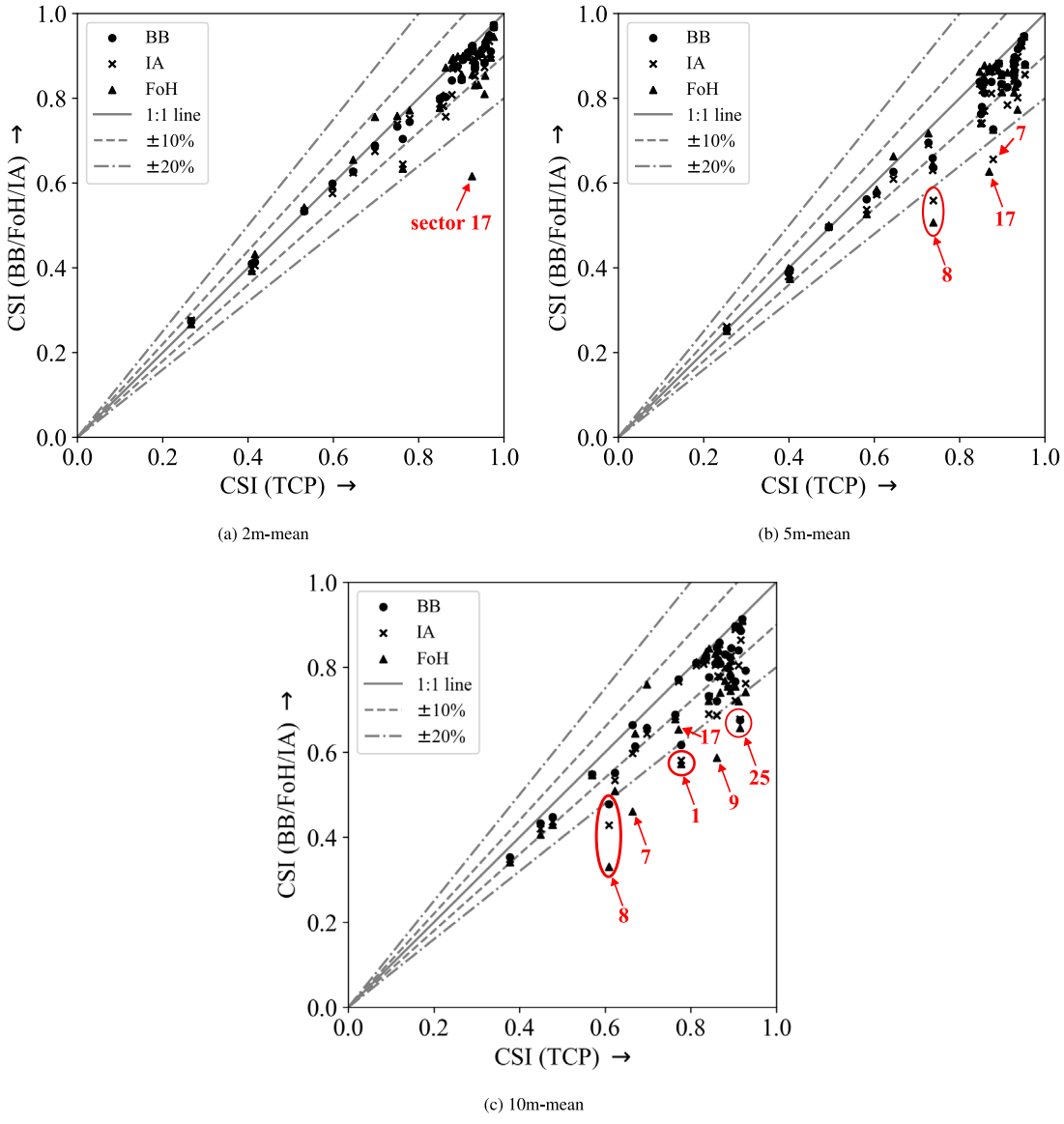


Fig. 3. Illustration of the methods of flood map reconciliation. Shading represents inundated area and vice-versa.



(a) 2m-mean

(b) 5m-mean

(c) 10m-mean

Fig. 4. Scatter plot of CSIs obtained using different reconciliation methods (for the use of mean resampling operator).

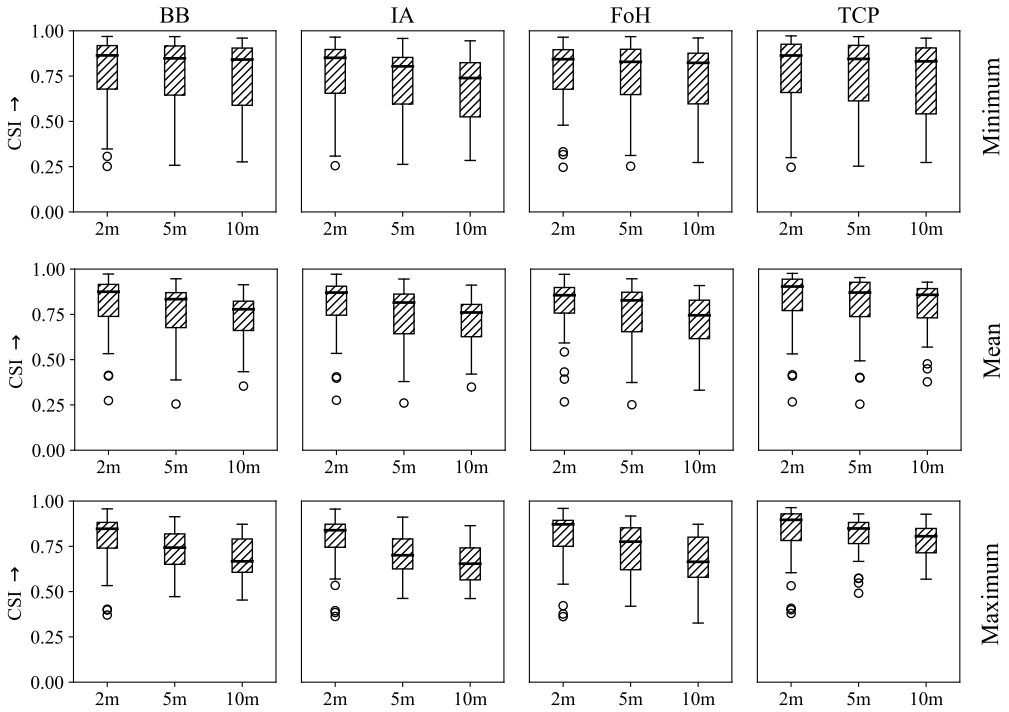


Fig. 5. Box and whisker plot of CSI for different model grid resolutions and DEM resampling operators.

An alternative method for flood map reconciliation is using a **terrain-constrained propagation (TCP) method**. It is a two step method which begins with the generation of a mask layer representing building footprints by taking the difference of the DEM and DTM. This mask is utilised to restrict computation to built-up areas only.

Thereafter, starting from the raw modelled WSE raster, the WSE is propagated through the built up areas. This propagation is contingent on the condition that the WSE remains above the DTM, ensuring that holes caused by actual topographic features are preserved while addressing the discrepancies due to buildings in the DEM.

The TCP method implemented herein is based on a numerical solution to the Eikonal equation using the Fast marching algorithm (Sethian, 1999):

$$|\nabla T(x)| = 1, \quad \boldsymbol{x} \in \Omega \tag{1}$$

where $T(x)$ represents the shortest-path distance from the known flood front to position $\boldsymbol{x}$, and Ω is the domain of solution.

The right-hand side of Eq. (1) being equal to 1 represents isotropic propagation. This implies that WSE values are interpolated based on the values at nearest (possibly multiple) upwind (already computed) cells.

2.3.2. Flood extent validation

Comparison of only the magnitude of inundated areas across different areal units to evaluate the performance of hydrodynamic models has a significant limitation. This is because it does not consider the spatial distribution of inundation within these areal units. As a result, models with similar total flooded areas but differing spatial distributions may yield misleading validation outcomes.

To overcome this limitation, more robust spatial validation metrics, such as the **critical success index (CSI)**, can be employed (Bennett et al., 2013). The CSI captures both the magnitude of inundated area and its spatial distribution by evaluating the overlap between observed and modelled inundation areas, considering agreement and discrepancies in each pixel's inundation status. It is defined as follows:

$$\mathrm{CSI} = \frac{A_\mathrm{h}}{A_\mathrm{h} + A_\mathrm{f} + A_\mathrm{m}} \tag{2}$$

where A_h represents 'hits' - the area correctly predicted as flooded in the modelled map, A_m represents 'misses' - the area flooded in the observed map but not in the modelled map, and A_f represents 'false-alarms' - the area flooded in the modelled map but not in the observed map.

Figure S3 illustrates how for the same magnitude of flooded area the CSI penalises different levels of spatial disagreement.

Additionally, since broad-scale evaluations may obscure localised discrepancies, the CSI should ideally be computed at smaller spatial scales rather than over large administrative units like municipalities. 28 smaller 'evaluation sectors' (hereafter referred to as 'sectors') are thus defined (Fig. A.2), facilitating localised performance assessments, helping to identify specific areas of model weakness. Also, Fig. A.3 shows the number of observation points in each sector, which may serve as a proxy for confidence in the flood extent raster in a given region.

2.3.3. Maximum flood depths validation

Two different approaches, utilising two different products, have been adopted to assess the model performance with respect to maximum flood depth: (a) by using the discrete set of points with associated water depths (Fig. 2) and (b) by using the continuous (IDW) raster of water depths (Figure S1).

Discrete points based validation: In this approach the water depth associated with each measurement point is converted into a water surface elevation (WSE) value by adding to it the underlying terrain elevation. This transformation is essential because WSEs provide a more stable and meaningful basis for comparing inundation patterns across regions with varying topography.

However, doing the same for the corresponding points in the model is not straightforward. A key shortcoming of the field survey data collection process is that the georeferencing of watermarks does not precisely match their exact position of recording. As a result, the recorded points are predominantly located 'within' building footprints, since they are derived from the structures where measurements were taken (Fig. 2). However, in the building-block BTM, hydrodynamic models typically do not assign water depths to such locations. Consequently, direct comparisons at these discrete points are infeasible.

Therefore, the TCP method is used to interpolate or extrapolate the WSE through the building footprints. This enables direct comparison at the location of discrete HWMs.

Continuous raster based validation: In this approach a continuous WSE raster is created from the observed and modelled water depth rasters by adding the terrain elevations. Thereafter, errors in WSEs are quantified by computing the difference between the observed and modelled WSE rasters.

As stated before, the observed water depth raster is based on the 2013–14 DTM, which reflects the best available product at the time of creation of the 'observed' products. In contrast, the modelled water-depths are derived using the 2021–22 DEM (representing the topographical conditions with closest temporal proximity to the time of flood) as input to the hydrodynamic model. To ensure methodological consistency, the observed WSEs are constructed using the 2013–14 DEM, while the modelled WSEs utilise the 2021–22 DEM.

$$\begin{aligned} & E_h = Z_M - Z_O \\ \Rightarrow\ & E_h = (h_M + D_M) - (h_O + D_O) \\ \Rightarrow\ & E_h = (h_M - h_O) + (D_M - D_O) \end{aligned} \tag{3}$$

where E_h is the error in water surface elevation, Z_M and Z_O are the modelled and observed water surface elevations, h_M and h_O are the modelled and observed water depths, and D_M and D_O represent the DEMs (topographic elevations) used for the modelled and observed datasets.

Smith et al. (2021) highlight the significant challenge linked with validation of hydrodynamic models against observed depths/HWMs owing to the intricate uncertainties associated with DEM resolution, resampling or even the exact nature of the products used. Wing et al. (2021) further highlight that certain metrics such as the root mean squared error (RMSE) are sensitive to outliers, and thus may give unrepresentative outcomes. The authors use the mean absolute error (MAE) in their study. Given this range of discussion, in the present study, different error metrics at various scales are computed in order to meticulously analyse model performance. A Bias metric (Eq. (4)) is employed to have a first view of the over- or under-predicting behaviour of the model and then domain-level MAE and RMSE values are computed as well (Equation (5) and (6)) (Wing et al., 2021). The final level of analysis comprises a detailed view of intra-sectorial variations using box-and-whisker plots of E_h, for all sectors.

$$\begin{aligned} \text{Bias} &= \frac{\sum_{i=1}^{N}(Z_M - Z_O)}{N} \\ &= \overline{E}_h \end{aligned} \tag{4}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N}|Z_M - Z_O|}{N} \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(Z_M - Z_O)^2}{N}} \tag{6}$$

where N is either the number of observed HWMs or the number of inundated pixels in the observed raster.

3. Results

3.1. Flood extent validation

When operating at the highest resolution, the model demonstrates consistently high performance across most sectors, achieving a peak sectorial CSI score of 0.97 and sectorial mean of 0.83 (using the TCP reconciliation method) (as shown in Fig. A.4(a)). On the other hand, sectors 30–32 (along the Hoëgne) and sectors 33–35 (along the Wayai) show contrastingly low CSI values (≤ 0.65). Agreement of modelled and observed flood extents is depicted in Fig. 4. While the TCP method emerges as the best performer, the four methods of reconciliation, across sectors, produce outcomes that are, on average, within ±5% of each other in terms of CSI (for case 2m-mean). However, there are notable instances viz. sector 1, 7, 8, 9, 17, and 25, as highlighted in Fig. 4, where some methods result in a (>) 20% poorer CSI as compared to that of the TCP method (see also, Fig. A.4). These deviations are discussed in Section 4.1.1.

In Fig. 5, each row corresponds to one of the three resampling operators, while each column depicts the use of one of the four reconciliation methods. Each individual plot conveys the variation of CSIs across sectors for the three grid resolutions. Increasing grid resolution usually results in improved CSI values. This improvement is found to be more pronounced when using the *mean* and *maximum* resampling operators. The *minimum* operator yields the most stable CSI values across all four reconciliation methods and resolutions although for the IA method the usual trend of declining scores for reduced resolutions is still observed. The sector-specific effect of each reconciliation method is however not reflected in Fig. 5, since it only depicts the spread of sectorial CSI across the entire domain for each method. For this, one must again turn to an explicit comparison plot (such as the aforementioned Fig. A.4).

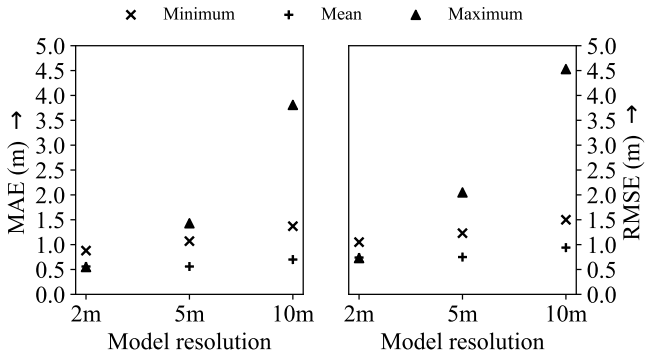


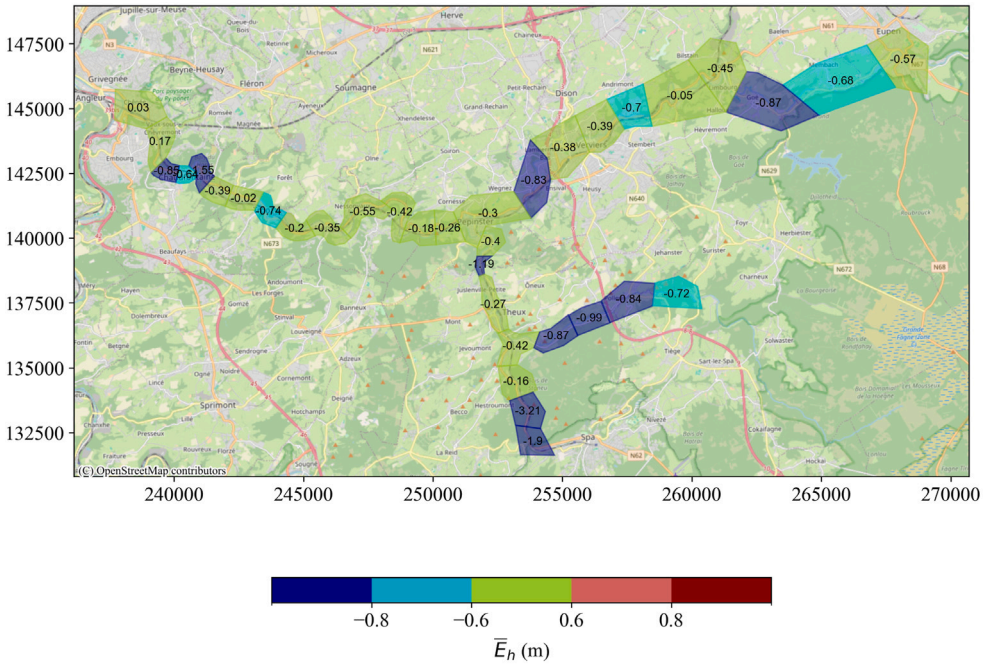
Fig. 6. MAE and RMSE in maximum flood depth prediction (for point-based evaluation).

3.2. Maximum flood depth validation

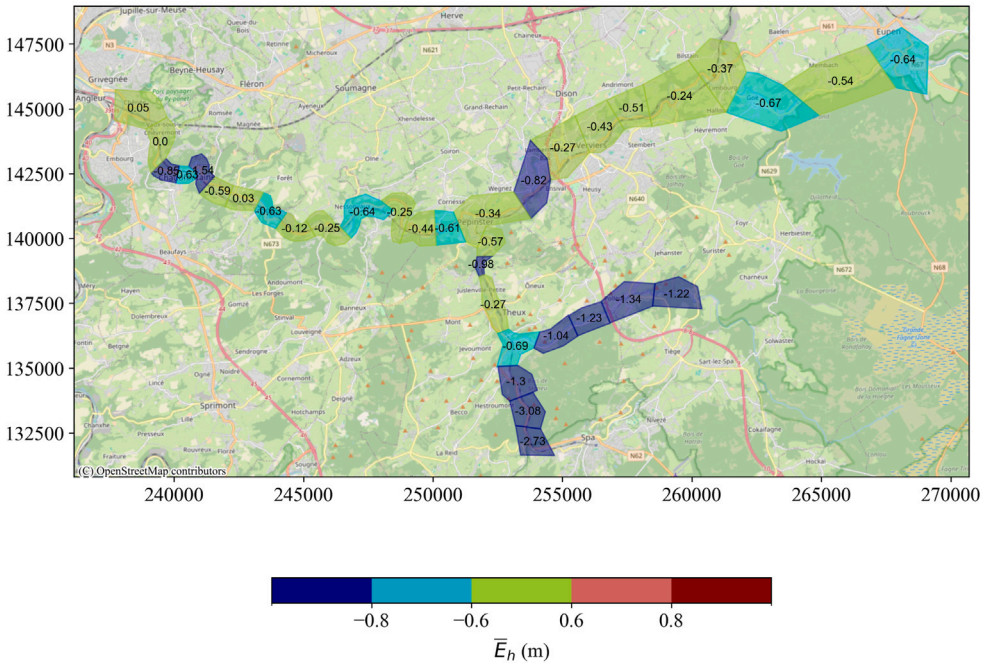
Fig. 6 depicts the domain-level MAE and RMSE values for each simulation. For the case 2m-mean, the MAE and RMSE are 0.56 and 0.74 m respectively. The errors rise with decreasing resolution. The resampling operator has an even stronger impact with errors rising with the tendency - mean < min < max, except for the highest resolution where the maximum operator yields a marginally improved MAE as the mean while the minimum operator model has higher errors.

Sector-wise $\overline{E}_h$ for the 2m-mean simulation are presented in Figs. 7(a) and 7(b) (for point- and raster-based comparisons, respectively). In most sectors along the Vesdre river the bias is between ±0.6 m. Four sectors along the Vesdre, viz. sector 3, 9, 21 and 23 show a strong underprediction bias while some others have marginal underprediction biases (sectors 2, 6, 18 and 22 for point-based comparison and sectors 1, 3, 12, 15, 18 and 22 for the raster-based comparison). On the other hand, along the Hoëgne and Wayai, 6 of the 11 sectors (for point-based comparison) presented strong negative mean bias errors.

Fig. 8 further shows the intra-sectorial variation of E_h (point-based comparison) across all evaluation sectors, for the use of the mean resampling operator. Outcomes corresponding to the other resampling operators are provided in Figs. A.8–A.10. Once again, as the computational mesh coarsens, an increase in errors is noticed although the median errors for the 2m-mean and 5m-mean simulations remain comparable.



(a) Point-based



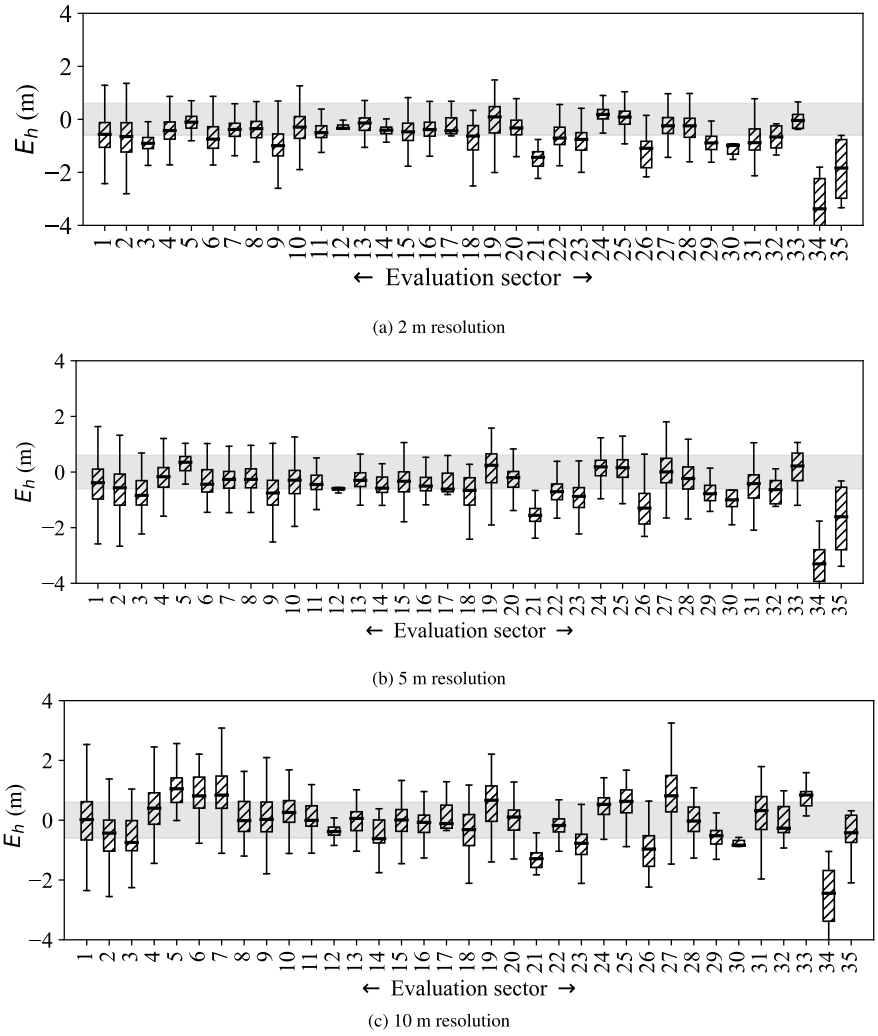
(b) Raster-based

Fig. 7. Sector-wise bias error in maximum flood depth estimation (2m-mean).

4. Discussion

4.1. Interpretation of validation outcomes

Fig. 9 combines the domain-level outcomes of flood extents and maximum flood depth validation to depict the overall performance of the 9 simulations. In Fig. 9, the vertical axis is inverted such that points (simulations) appearing further towards the top-right of the plot show better overall performance. Here, one finds two consistent trends. Firstly, grid refinement (10 m (red) to 5 m (blue) to 2 m (green), in Fig. 9) leads to improvement of both extent and depth agreements. This improvement is increasingly substantial when going from a mean resampling operator to minimum and maximum. Secondly, for a given resolution, the extent agreement is consistently better for the mean resampling case than the maximum (than the minimum). On the other hand, the maximum flood depth agreement follows the trend - mean, minimum and maximum (in order of increasing errors). However, at the highest resolution the maximum operator simulation (MAE = 0.55 m) performs comparably to the mean operator simulation (MAE = 0.56 m). In terms of CSI, the 2m-mean simulation has a score of 0.86 compared to 0.84 for simulation 2m-max. Nevertheless, from Fig. 9 it is clear that the errors yielded by the maximum operator simulations



(a) 2 m resolution

(b) 5 m resolution

(c) 10 m resolution

Fig. 8. Box and whisker plot of WSE difference (point-based comparison) for different model grid resolutions and *mean* DEM resampling operator.

are significantly more sensitive to model resolution than those using the mean operator. Therefore, the 2m-mean simulation is deemed to be the 'reference/best-case' in the present study.

4.1.1. Flood extents

On average, sectorial performance reaches approximately 0.83. The domain-level CSI of 0.86 is noteworthy when compared to values reported in the literature (de Arruda Gomes et al., 2021) (Table 1). Large-scale studies which rely on field data for validation have not reported such values. Studies that evaluate model outcomes against other benchmark numerical models have reported higher CSI scores (Mosimann et al., 2024; Falter et al., 2013). However, these comparisons often reflect consistency between models rather than absolute accuracy, as both models, even if differing in formulations, may share common underlying assumptions, parameterisations, or simplifications.

As stated in the previous section, the TCP reconciliation method consistently outperforms the others across most sectors (Fig. 4). Additionally, in some sectors (as indicated in Fig. 4), certain methods are seen to produce a markedly different CSI score than the others. Upon examination of such sectors further insights can be drawn.

For instance, Fig. 10 shows the inundation map for sector 8 which is representative of a highly urbanised unit. Here, progression of flood waters is increasingly restricted along different road networks as the model resolution coarsens. Fig. A.11 illustrates the declining accuracy of street network resolution with the coarsening of model grid, thus affecting conveyance. This, in-turn, leads to urban blocks that are not completely enclosed by flood waters. In such cases, the TCP method is able to propagate the water surface through the buildings till it is stopped by the terrain itself, while the BB approach manages to

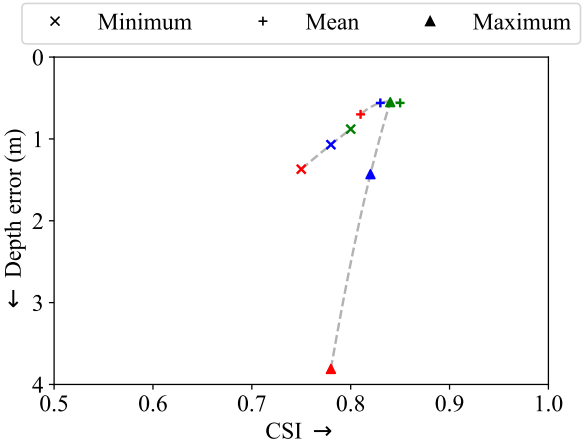
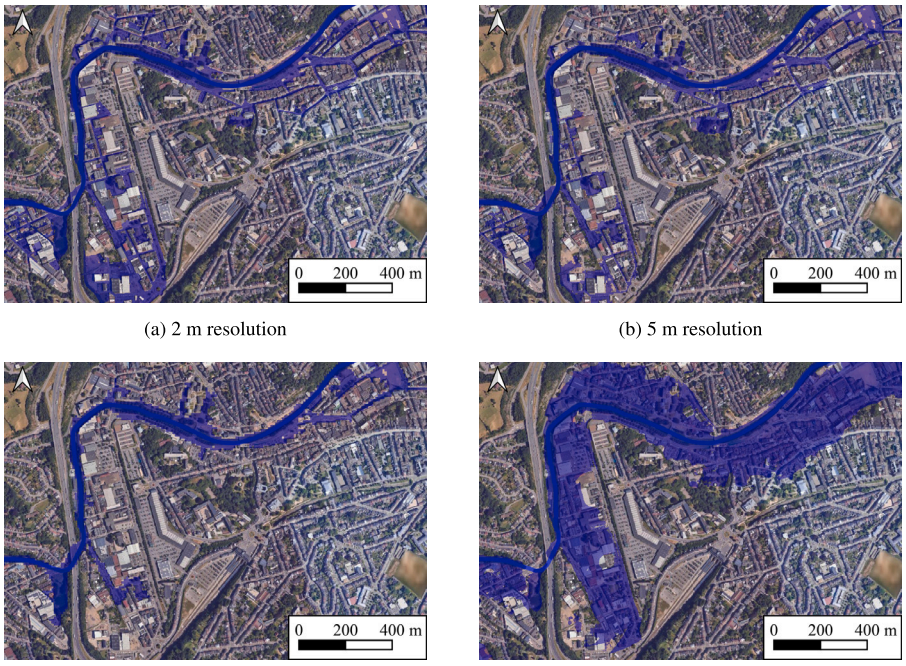


Fig. 9. Model performance with respect to flood extents and maximum flood depth. Note: green, blue and red symbols indicate 2 m, 5 m, and 10 m grid resolutions respectively.



(a) 2 m resolution (b) 5 m resolution

(c) 10 m resolution (d) Observed inundation

Fig. 10. Inundation maps of a highly urbanised area (evaluation sector 8) - based on (a) 2 m, (b) 5 m, (c) 10 m resolution models and, (d) field data.

identify some buildings of the urban block as inundated (those whose building polygons encounter an inundated cell in its vicinity). The IA approach attempts agreement with the observed inundation by removing building footprints from the observed inundation map. For the FoH approach, however, this represents a significant limitation, as the method does not recognise several inundated built-up areas since it does not find 'holes' to fill, which leads to significant underprediction (Fig. A.12) and thus a poor CSI score as reflected in Fig. 4. The same is the case for sector 1, 7, 9 and 25 which are similarly urbanised. In any case, at lower resolutions even the IA (already at 5 m resolution) and BB methods (at 10 m resolution) fail to reconcile the differences between the raw modelled and observed flood maps.

For sector 17 (Fig. 11), visual inspection suggests that the modelled inundation rasters for the 2 m and 5 m models are in very good agreement with the observations. The 10 m resolution model overpredicts slightly. While the TCP, BB and IA methods would function without fallacies in such a scenario (i.e., make minor interventions at the few buildings), the FoH method ends up filling region 1 and region 2 (since they are 'holes') (Fig. A.13) which in-turn lowers the CSI score. Notably, for region 2 a minor correction was made in the DEM in the form of introducing a railway underpass so that the conveyance is not artificially blocked in the hydrodynamic model (see Fig. A.14). Paradoxically, had the correction not been made region 2 would cease to be a hole and therefore the CSI would not be penalised as much.

The instances discussed above relate to the outcomes of the mean resampling operator simulations at various resolutions. It has been further observed that the minimum resampling operator poses problems at coarser resolutions particularly to the IA method. This may be illustrated using the example of Sector 27 (Fig. 12). Here, lower-rise structures and isolated built-up areas are increasingly flattened out due to resampling to reduced resolutions. This shows itself in the simulated inundation maps at different resolutions.

At 2 m resolution (Fig. 12(a)), most buildings appear as holes in the inundation map. As the resolution decreases, fewer buildings retain their elevation footprints in the DEM and therefore floodwaters are found to flow over them.

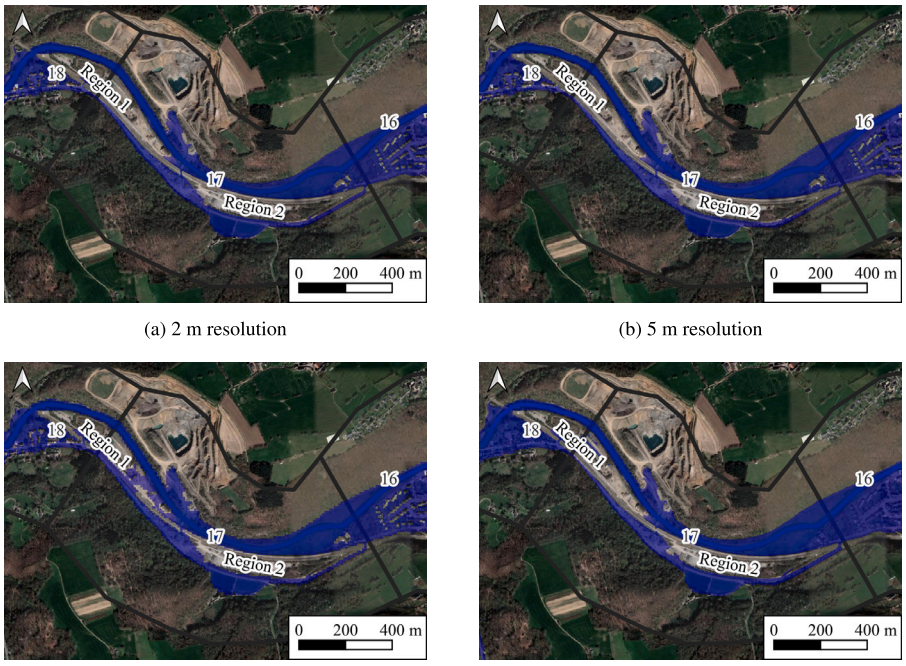
For the BB, FoH and TCP reconciliation methods, this does not present significant concerns because the three approaches involve adding building footprints to the computed inundation map. At lower resolutions, if the building footprints are already inundated (as evident in Fig. 12(c)), these methods do not introduce any changes. However, this situation poses a challenge for the IA method. The IA method would incorrectly remove several building footprints from the observed inundation maps (Fig. A.15), resulting in poorer CSI scores when comparing with the lower-resolution flood maps. Fig. A.16, which compares the four reconciliation methods for the 10m-min simulation, depicts the recurrence of this problem of the IA method across various sectors of the simulation domain.

4.1.2. Maximum flood depth

For the reference model (2m-mean), the MAE of 0.56 m is very acceptable and comparable to others found in literature (Table 1). Further, Fig. 13 shows that the highest errors are encountered in sectors with less than 2% of the total observation points. In fact, sectors 34 and 35, which yield very strong underprediction errors, contain less than 0.07% of all observation points. In the same way, low errors in other sectors with such minority of observation points are not viewed (relatively) as highly reliable markers of good performance. Beyond the 2% mark, the errors are well-restricted to the ± 0.6 m bracket, in all but one sector.

In Fig. 14, it is found that in all but one sector, the sectorial bias errors computed based on points and those based on rasters, are within ± 0.6 m of each other. This may be dependent on the validity of the IDW approach in specific sectors, among other factors.

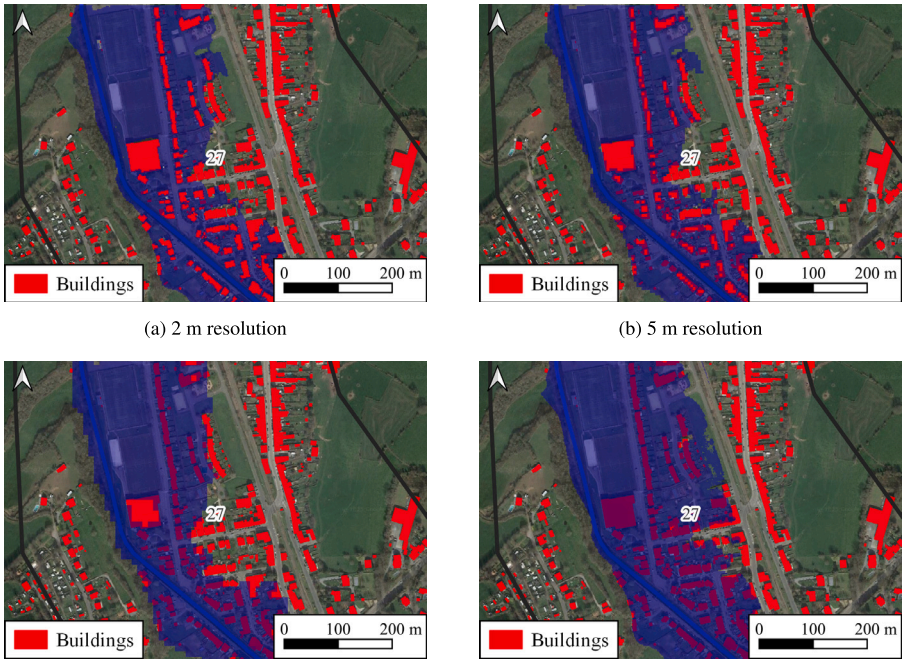
Further, Eq. (3) decomposes the observed and modelled WSEs into their respective water depth and DEM components. This separates the error into two components: $h_M - h_O$, which quantifies the difference in water depths, and $D_M - D_O$, which accounts for differences in the DEMs used. Ideally, the latter should be equal to 0. However, this is not always the case, as in the present case where different versions of DEMs were utilised for the model and the observed raster. Therefore, this could add on to (or compensate for) the actual error in water depths depending on how the DEM affects the flow dynamics. The discrepancy in the topographical layer versions used for observed and modelled WSE calculations is likely to have a greater influence on the continuous raster-based evaluation.



(a) 2 m resolution (b) 5 m resolution

(c) 10 m resolution (d) Observed inundation

Fig. 11. Inundation maps of a non-urbanised area (sector 17) - based on (a) 2 m, (b) 5 m, (c) 10 m resolution models and, (d) field data. Note: Black lines indicate evaluation sector boundaries.



(a) 2 m resolution (b) 5 m resolution

(c) 10 m resolution (d) Observed inundation

Fig. 12. Inundation maps of sector 27 - based on (a) 2 m, (b) 5 m, (c) 10 m resolution models and, (d) field data.

Nevertheless, maximum flood depth outcomes may largely be considered a function of (a) model grid resolution, and (b) DEM resampling operator.

As reported in several past studies (Muthusamy et al., 2021; Shen and Tan, 2020; Savage et al., 2016), it is found that for a given resampling operator, errors tend to increase significantly with decreasing model resolution. While the higher error across sectors (Fig. 8) is not observed to be biased in either direction, in urbanised sectors (such as sectors 5, 6, 7 or 27) grid coarsening is found to predominantly increase predicted water depths. This is in line with the findings of Muthusamy et al. (2021).

The choice of DEM resampling operator has a much stronger influence on maximum flood depths. Across Figs. A.8–A.10, the minimum and maximum operators are found to introduce clear biases in maximum flood depths (negative and positive, respectively). The extent of bias introduced increases with decreasing resolution. Further, it is found that at the domain level the use of the maximum operator gives 1.6 and 3 times higher root-mean-squared errors than the minimum

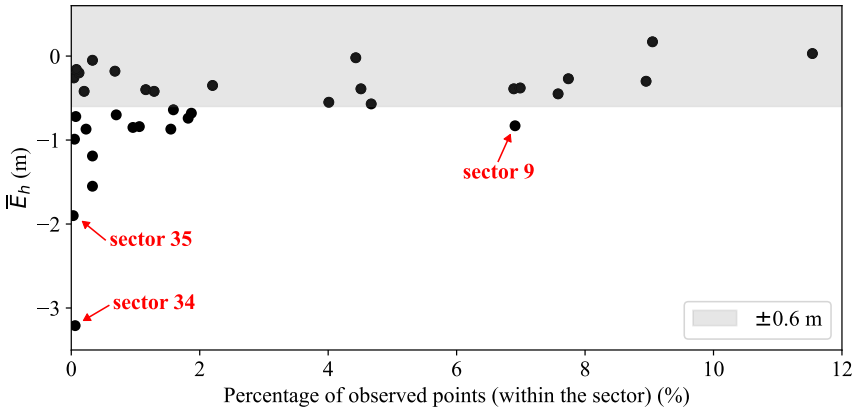


Fig. 13. Sector-wise mean bias error in maximum flood depth prediction as a function of the percentage of total observation points (2m-mean).

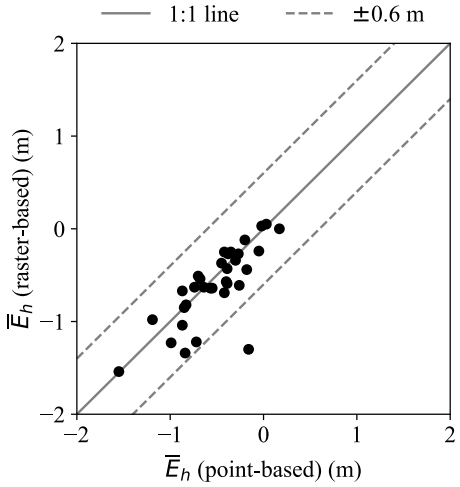


Fig. 14. Comparison of point- and raster-based sectorial bias errors (2m-mean).

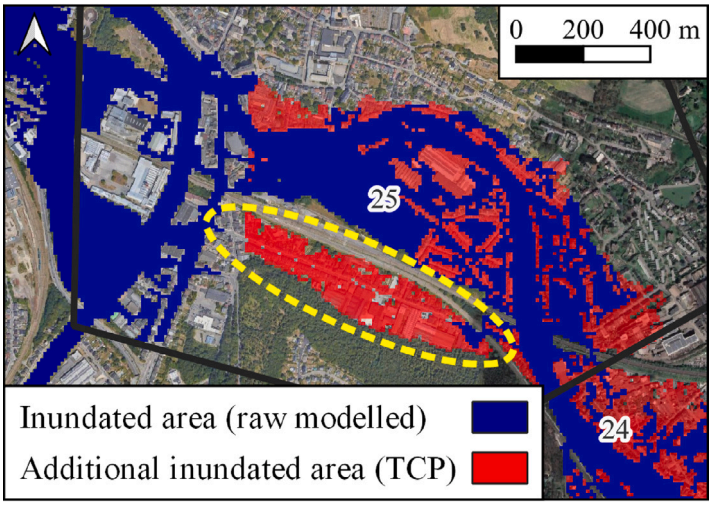


Fig. 15. Example of a potential over-compensating effect of the TCP method (sector 25).

operator for the 5 m and 10 m resolution cases, respectively (see also, Figs. A.9a and A.9c, or Figs. A.10a and A.10c). This could be attributed to the fact that the minimum operator retains conveyance capacities in areas such as urban districts (streets), narrow reaches of rivers, among others. The maximum operator, on the other hand, is much more likely to impede flood propagation which would lead to large localised errors as well as potential propagation via back-water effects (Muthusamy et al., 2021).

Finally, perusing Fig. 9 highlights the importance of validating inundation models against both extents and depths. This is advocated strongly by the simulation 5m-max, where there is a very high-degree of agreement in terms of flood extents (CSI = 0.82) but very strong overprediction bias for maximum flood depths.

4.2. Flood map reconciliation method

The need for reconciliation of observed and modelled flood maps, when modelling based on the building-blocks BTM, has been detailed in former sections. However, the choice between the BB, IA, FoH and TCP approaches involves a careful consideration of the characteristics of the study area, the purpose of the validation, data availability and the specific limitations of each method that merit detailed evaluation.

The BB approach provides the benefit of assisting in highly urbanised areas while not meddling with outcomes in non-built-up environments. It allows for contextual flexibility by virtue of the variable buffer size. This adaptability would enable users to fine-tune the method based on specific study conditions. In the present case, the domain-level CSI is found to not be sensitive to buffer radii varying from one to two times the cell dimensions (CSI = 0.82 in either case, for 2m-mean). But caution should be maintained and sensitivity to buffer sizes ought to be checked. Determining the appropriate buffer size depends on the complexity of urban block structures and model resolution. The former could additionally vary between regions in a large-scale simulation. Regardless of its effect on the CSI, overly large buffers could lead to significant overestimation of the number of inundated buildings (see for example, Figure S5), which is often a metric of importance in flood reports.

The IA approach is similar to the BB approach in that it restricts modifications to areas where buildings exist. While theoretically the IA approach might appear to be the most logically sound, there are certain drawbacks and pitfalls to look out for. While this method avoids overestimating flooded areas in the manner the BB approach might, it may inaccurately remove inundated areas from the observed flood map. This issue becomes pronounced during extreme flooding events with several meters of water depth, where low-rise structures might genuinely be inundated. This aspect of the IA approach conflicts with the building-block BTM where buildings can be inundated under sufficiently high water levels (unlike the 'building hole' approach, which treats building footprints as areas of non-computation). Also, as illustrated in Fig. 12, this effect is increasingly substantial at reduced mesh resolutions and upon the use of the minimum resampling operator. Finally, it is a significant shortcoming of this method that the total inundation area computed would be lower than the real values, since the method excludes built-up areas. Therefore, this method only serves the purpose of enhancing comparability of observed and modelled rasters for the sake of validation and is not suitable when the goal is to report on the total inundated area due to a flood.

Both approaches rely on the availability of geospatial building datasets, which is not commonplace for several parts of the world. Even when available, it is wise to inquire about the period of record. An outdated dataset may miss major constructions while data recorded long after a flood event might comprise additional buildings that did not exist at the time of flood. Some instances of this are found in Figure S5, where the building-polygon dataset is seen to miss certain buildings visible in the satellite image.

The FoH approach has the merit of being the most straight forward and computationally efficient. It is particularly useful in regions with simple topography and building layouts. However, the hit-or-miss nature of the FoH approach has already been highlighted in Section 4.1.1. It carries inherent risks of both overestimating and underestimating flooded areas (e.g. Fig. 11). As a specific concern, models that have resolutions coarser than (or comparable to) the typical width of streets in a network are prone to inaccurate physical representation of the street networks. This would lead to inaccurate flood conveyance through the network (Fig. 10(c)), increasing the chances of non-enclosed inundation areas - which is a pitfall for the FoH method, resulting in underestimation of inundated area. Conversely, genuine dry patches, caused by elevated terrain rather than buildings, might be inaccurately filled, leading to overestimation.

The risk of incorrect filling is reduced by the TCP method. Once again, however, the applicability of this method is contingent on the availability of both the DTM and DEM for the simulated domain. The TCP method also comes with the benefit of being able to act in extrapolation. For the present case, this approach yields improvements greater than 5 percent for 11 different sectors when compared to the FoH method (Figure S6). The simplicity of the FoH method, its computational ease and applicability in data-scarce domains may yet prompt modellers to adopt it instead of the above 'improvement' which is significantly more computationally intensive.

While the TCP method demonstrates potential in mitigating resolution-induced problems in flood extents agreement, it also introduces the caveat of potentially overcompensating for legitimate physical limitations of the simulation. In doing so, it may inadvertently undermine meaningful model deficiencies such as those related to flow conveyance, resistance, or structural blockages. One such instance is depicted in Fig. 15 (sector 25 - simulation 10m-mean). Here, water in the raw model output appears to be unable to propagate through a cluster/chain of buildings (highlighted using the yellow dashed ellipse). Once again, this is attributable to the hindrance generated in street conveyance due to resampling. Nevertheless, the application of the TCP method here results in an extrapolation of the water surface through more than 500 m of the building chain. Such an enforcement of a smoothed water surface, that does not arise naturally from the governing physics of flow through urban street networks, may brink on the realm of 'bathtub modelling' as critiqued by Sanders et al. (2024). Thus, the application of the TCP method should be followed by an assessment of whether the resulting inundation pattern reflects expected physics or serves merely as a heuristic correction.

4.3. Post-flood surveys

HWMs, in principle, are discrete points (coordinates) with associated water depths/levels. Very frequently, they are recorded on building facades (Koenig et al., 2016), the present case being no exception. In the field survey data used here, while the association of a HWM with a building is visible upon visualisation, nothing is known about the precise geolocation of these points i.e., neither are they centralised in the building polygons nor do they show clear proximity to a particular facade (Fig. 2). A work-around for this issue is the TCP method as stated in Section 2.3.3. In case sufficient data is not available to execute the TCP method, a nearest-neighbour approach may be adopted to find the closest inundated point in the modelled raster to a discrete field observation point (illustrated in Figure S4). However, adopting the nearest-neighbour approach, for cases where there are strong variations in WSE around structures would introduce an uncertainty in maximum flood depth validation. In fact, this could be a factor that prompts the use of a continuous rasterised evaluation.

To address these challenges, future post-flood surveys can be strategically designed to collect data more compatible with relevant model practices. For HWMs on building facades, rather than allocating a water depth to the building as a whole, precise geolocation or indication of the particular facade where the HWM was recorded would reduce uncertainty in post-processing. Apart from building facades, surveys could also prioritise locations that allow for more direct comparison. For instance, it is beneficial to record HWMs at structures that experience similar water depths/levels all around it, such as trees, poles, and strategically placed crest-stage gauges (Koenig et al., 2016).

4.4. Error attribution

The quality, resolution and version of topographic data used influences both the observed and modelled datasets. For the former this directly affects the flood extents map and the reconstruction of WSEs, while for the latter it does the same by affecting the physics of the flow. In the aforementioned case-study, the regional authorities generated the raster product using topography data from 2013–14, which may no longer accurately reflect evolved conditions (at the time before or after the flood event) across the domain. Over time, natural and anthropogenic changes — such as erosion (which may be severe during mega-floods), construction, or vegetation growth — can alter the landscape. Such discrepancies introduce a layer of uncertainty when using the raster for model validation.

Additionally, the spatial distribution of the field data significantly impacts the reliability of the flood extents. In regions with sparse observations (Fig. A.3), interpolation techniques used to produce the 'observed raster' may fail to capture localised variations. Moreover, areas where the raster is extrapolated beyond the spatial bounds of the observation points are particularly vulnerable to inaccuracies. These factors not only impact the representativeness of validation metrics, such as the CSI, but also hinder true performance assessment of the hydrodynamic model in areas of sparse data, making it difficult to discern whether observed discrepancies in these areas stem from model inaccuracies or are artifacts of the observation dataset itself.

Disentangling these effects is non-trivial and requires careful consideration of model limitations, reconciliation methods and the quality of observed data. Therefore, modellers must maintain caution while attributing errors to any particular facet of the modelling exercise alone.

5. Conclusion

This study, aided by an unprecedented post-flood survey in the aftermath of the July 2021 mega-floods in Belgium, evaluates the performance of a detailed large-scale hydrodynamic model against flood extents and maximum flood depths. In examining the former, four flood map reconciliation methods - buffered-building (BB), inverse adaptation (IA), filling-of-holes (FoH) and terrain-constrained propagation (TCP) - are proposed, and their effectiveness in the validation process is tested.

The study highlights the importance of selecting reconciliation method based on the study area's characteristics and the intended purpose of the analysis. The BB method excels in urbanised areas, offering flexibility and context-sensitive adjustments but requires selection of an appropriate buffer size to avoid over- or under-estimation of flood extents. Conversely, the IA method ensures greater comparability in urban areas but reports an underestimate of total inundation due to its exclusion of built-up regions. The FoH method, while computationally efficient, is more prone to errors in elevated regions, complex urban terrains or where grid resolutions are insufficient to capture conveyance characteristics of features such as narrow streets. The TCP method performs the best in terms of reconciling differences but may sometimes tend to overcompensate model deficiencies as illustrated in Fig. 15. Future work could focus on the development of a systematic approach to dynamically apply these reconciliation methods based on the nature of terrain, urbanisation, model resolution etc.

The findings reveal strong agreement between modelled and observed outcomes, achieving a peak CSI score of 0.97 and an average of 0.83. However, its performance declines at coarser resolutions. The MAE in maximum flood depth estimation is found to be 0.56 m at 2 m resolution and when using the mean resampling operator. Additionally, the influence of DEM resampling operators on maximum flood depth predictions is significant, with the choice of operator introducing consistent biases. The mean operator provides the best compromise in this regard. Between the minimum and maximum operator, the former better retains conveyance features, while the latter, at coarser resolutions, tends to exacerbates localised errors by impeding flow propagation. Careful selection of resampling operators is therefore recommended, with respect to model resolution, particularly in urban contexts where terrain complexity affects model accuracy.

The analysis also highlights limitations in the post-flood survey process, particularly the placement of HWMs within building footprints, which complicates direct validation against model outputs. Recommendations are proposed to design post-flood surveys in a manner that facilitates model validation by reducing the highlighted uncertainties.

Ultimately, the study shows the challenges of detailed large-scale flood model validation, where uncertainties in observational data, terrain representation, and reconciliation methods can interfere with the accurate evaluation of model performance. Moving forward, the integration of updated and high-resolution geospatial datasets, as well as strategically designed validation datasets, will be critical in improving model reliability and applicability. Far from being the last piece of the puzzle, high-resolution observational datasets highlight great scope for further research in the domain of flood model validation.

CRediT authorship contribution statement

Pratik Chakraborty: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Christophe Dessers:** Software, Resources, Methodology, Data curation. **Pierre Archambeau:** Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Michel Pirotton:** Resources, Funding acquisition. **Sébastien Erpicum:** Resources, Funding acquisition. **Benjamin Dewals:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was funded in part by the Interreg project ResiRiver.

Appendix A

See Figs. A.1–A.16.

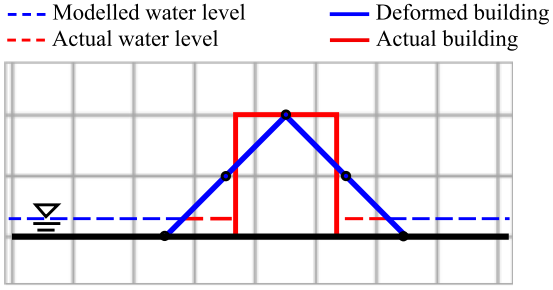


Fig. A.1. Deformation of buildings in digital elevation models (DEMs).

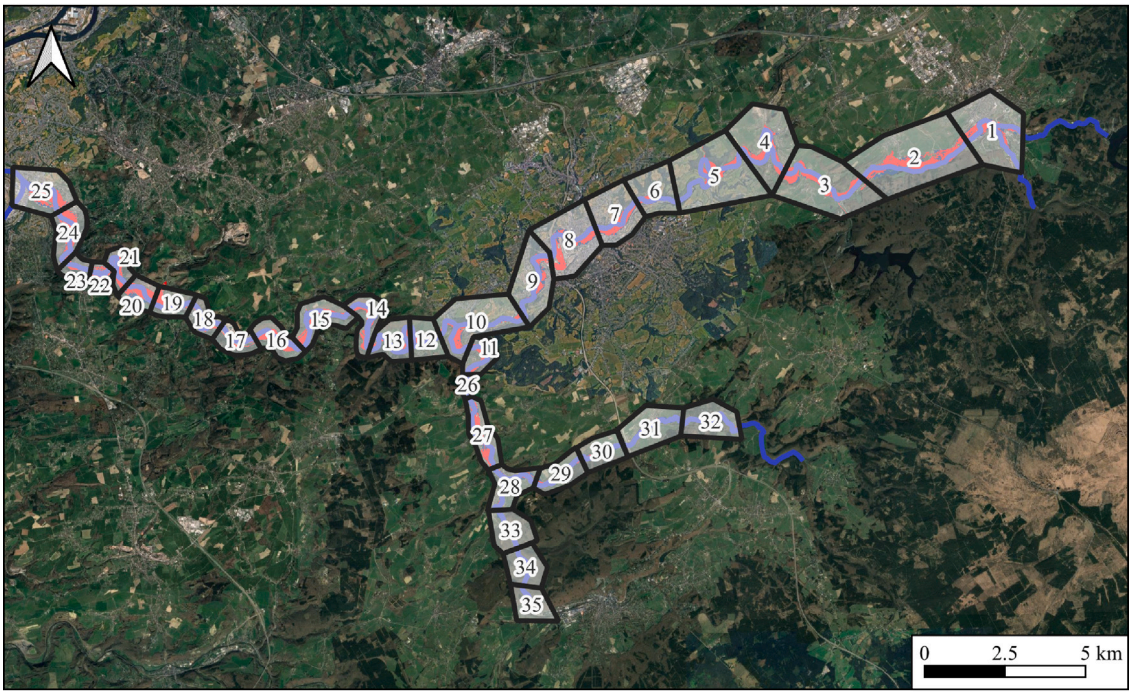


Fig. A.2. Areal units used for model evaluation.

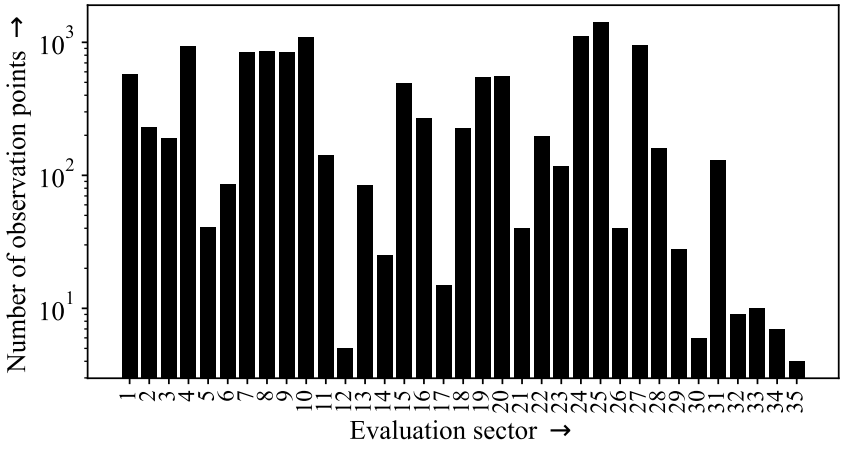
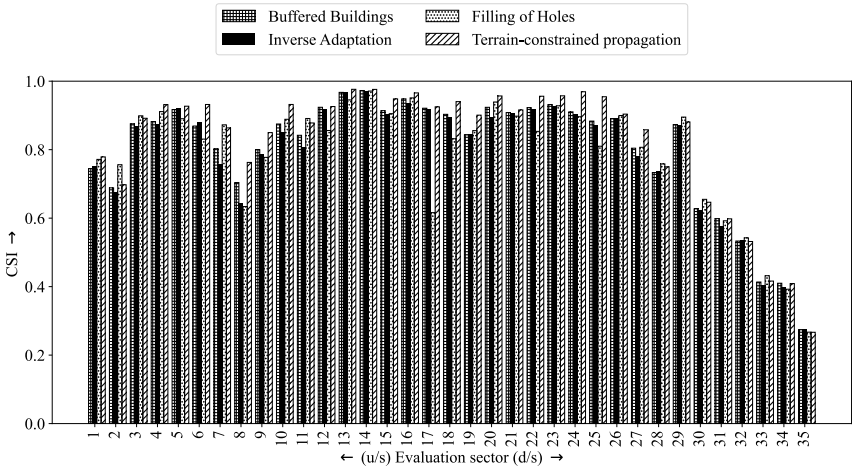
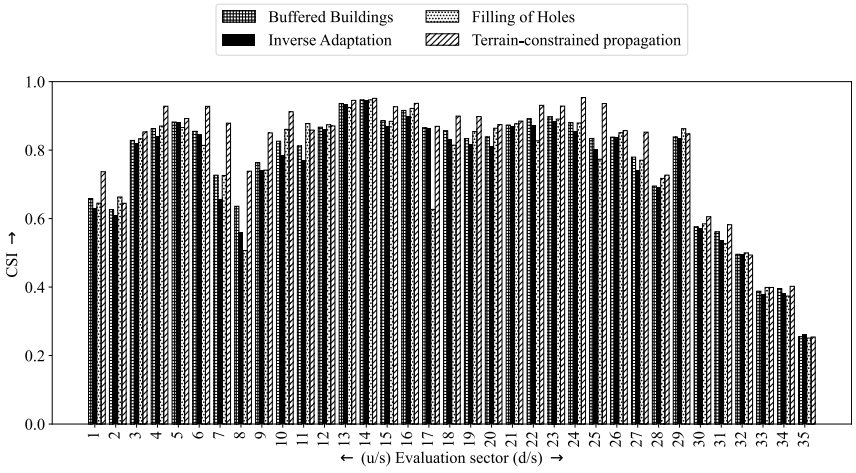


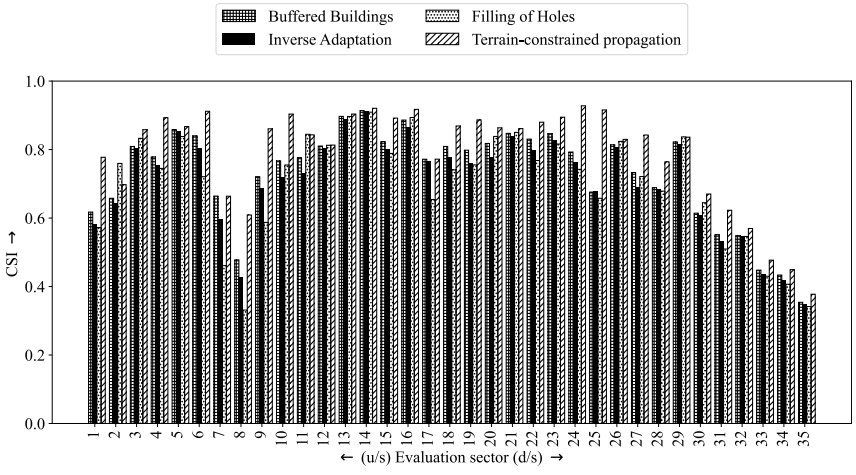
Fig. A.3. Number of observation points per sector.



(a) 2 m resolution

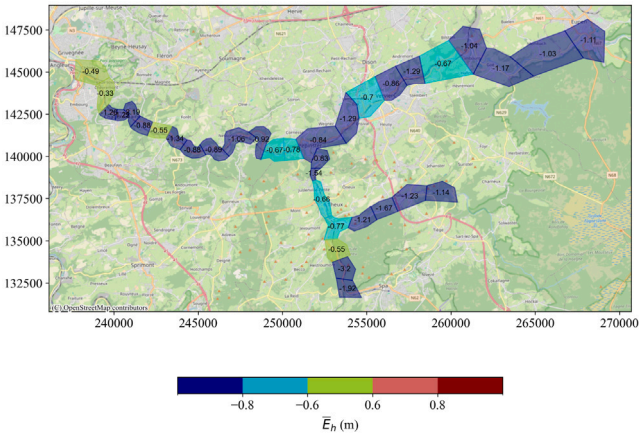


(b) 5 m resolution

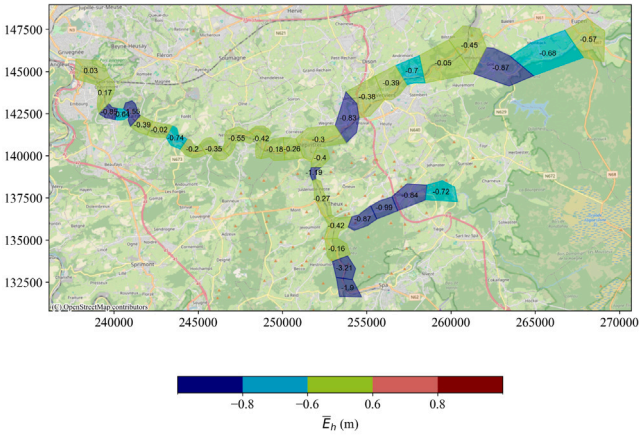


(c) 10 m resolution

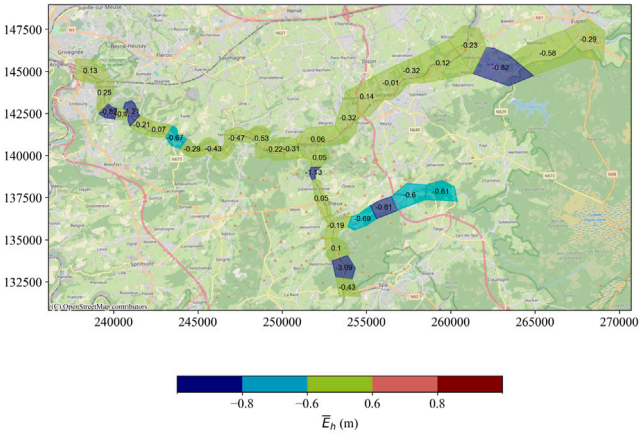
Fig. A.4. CSIs across evaluation sectors for simulations with different grid resolutions. Note: These results correspond to the use of the *mean* resampling operator.



(a) Minimum

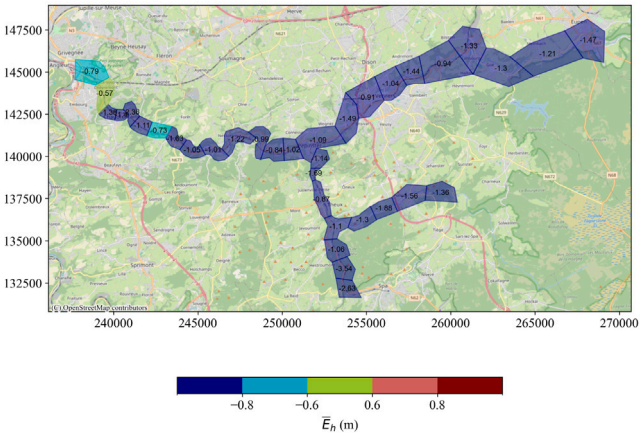


(b) Mean

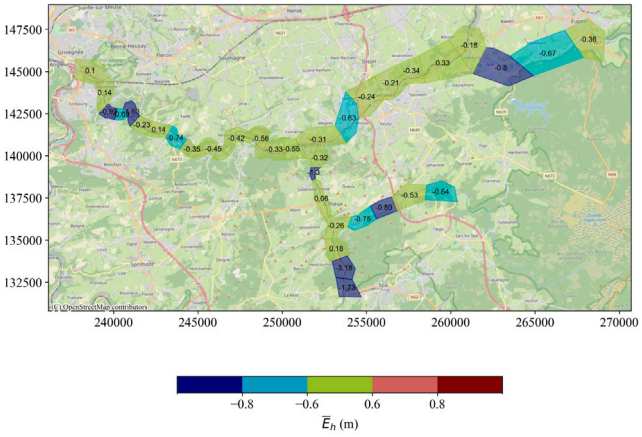


(c) Maximum

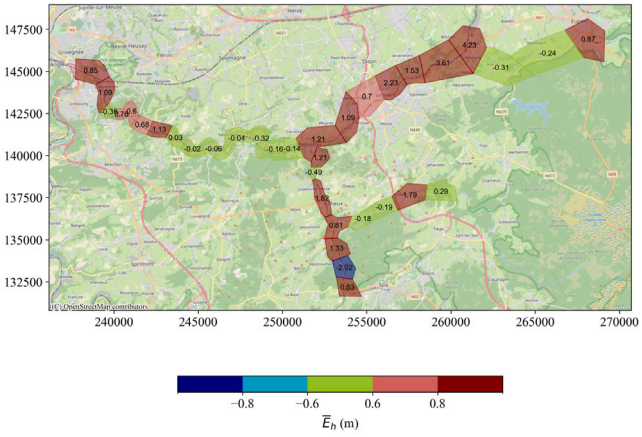
Fig. A.5. Sector-wise bias error in maximum flood depth estimation for 2 m grid resolution (point-based).



(a) Minimum

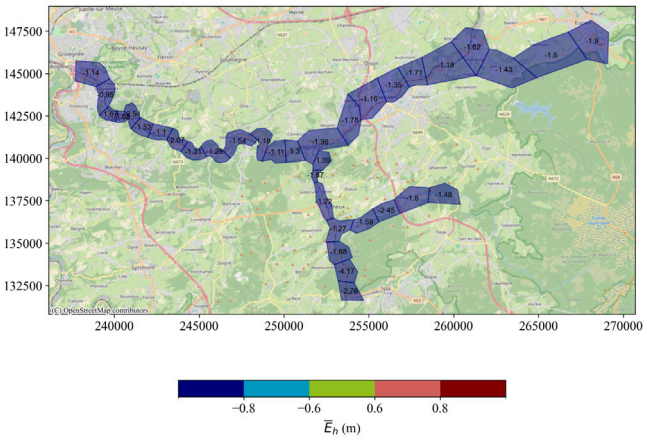


(b) Mean

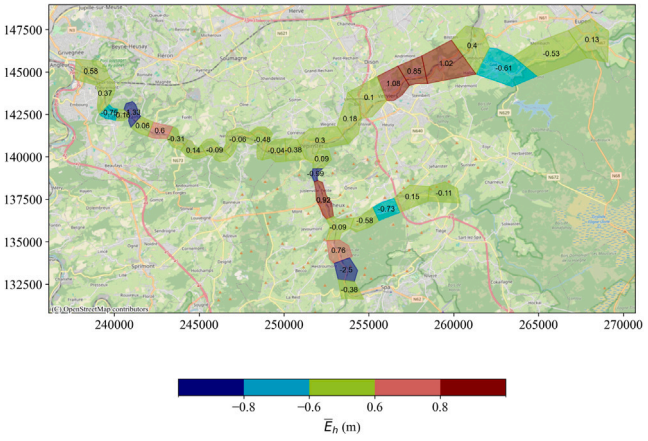


(c) Maximum

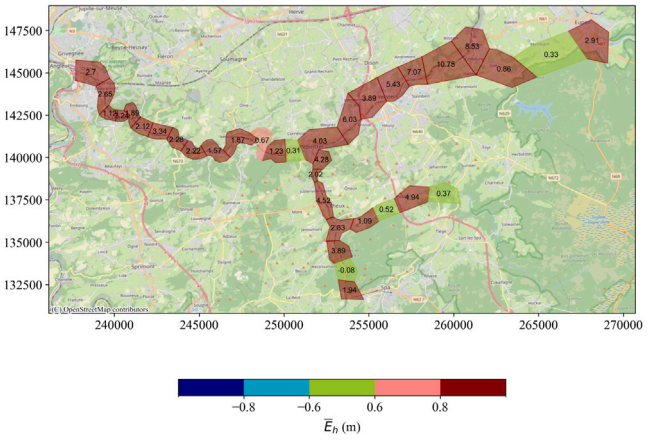
Fig. A.6. Sector-wise bias error in maximum flood depth estimation for 5 m grid resolution (point-based).



(a) Minimum

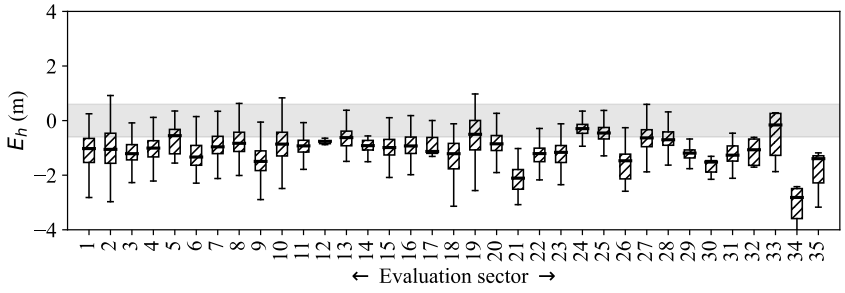


(b) Mean

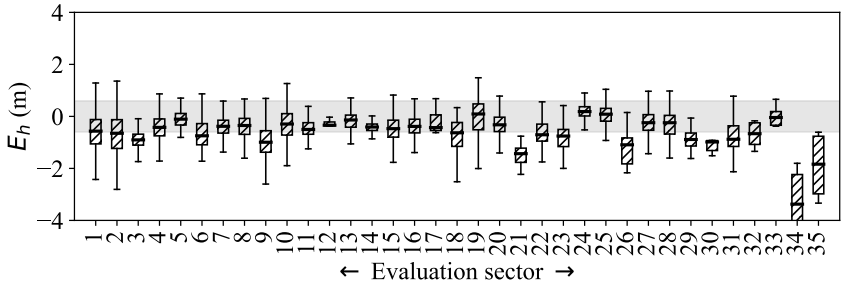


(c) Maximum

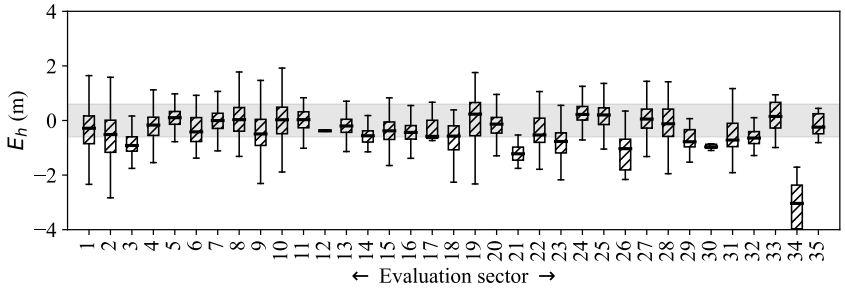
Fig. A.7. Sector-wise bias error in maximum flood depth estimation for 10 m grid resolution (point-based).



(a) Minimum

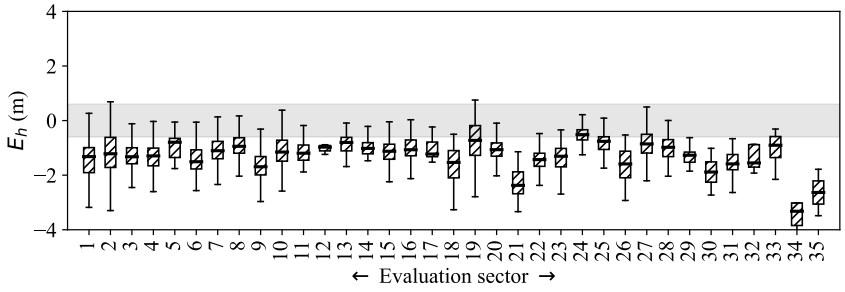


(b) Mean

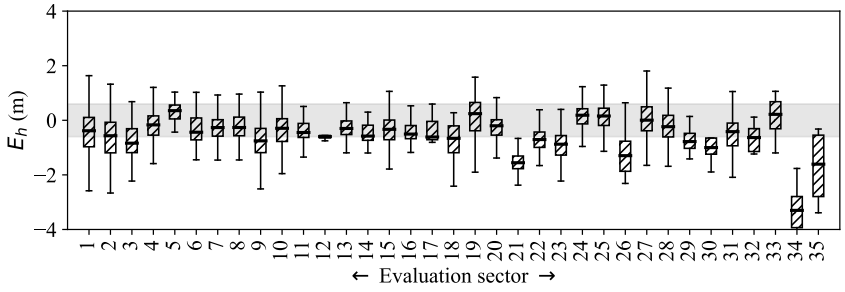


(c) Maximum

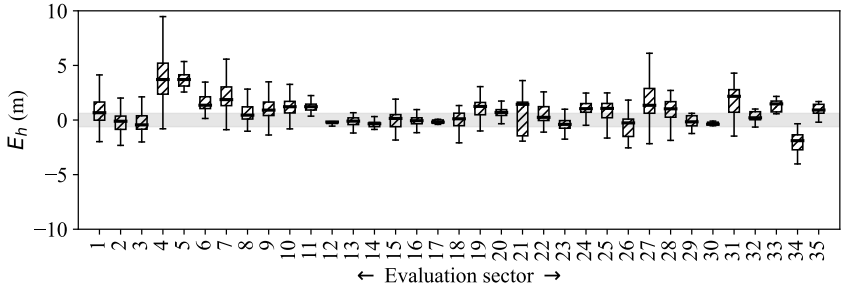
Fig. A.8. Box and whisker plot of WSE difference (point-based) for different resampling operators and 2 m grid resolution.



(a) Minimum



(b) Mean



(c) Maximum

Fig. A.9. Box and whisker plot of WSE difference (point-based) for different resampling operators and 5 m grid resolution. Note: Vertical scale is altered for sub-figure (c).

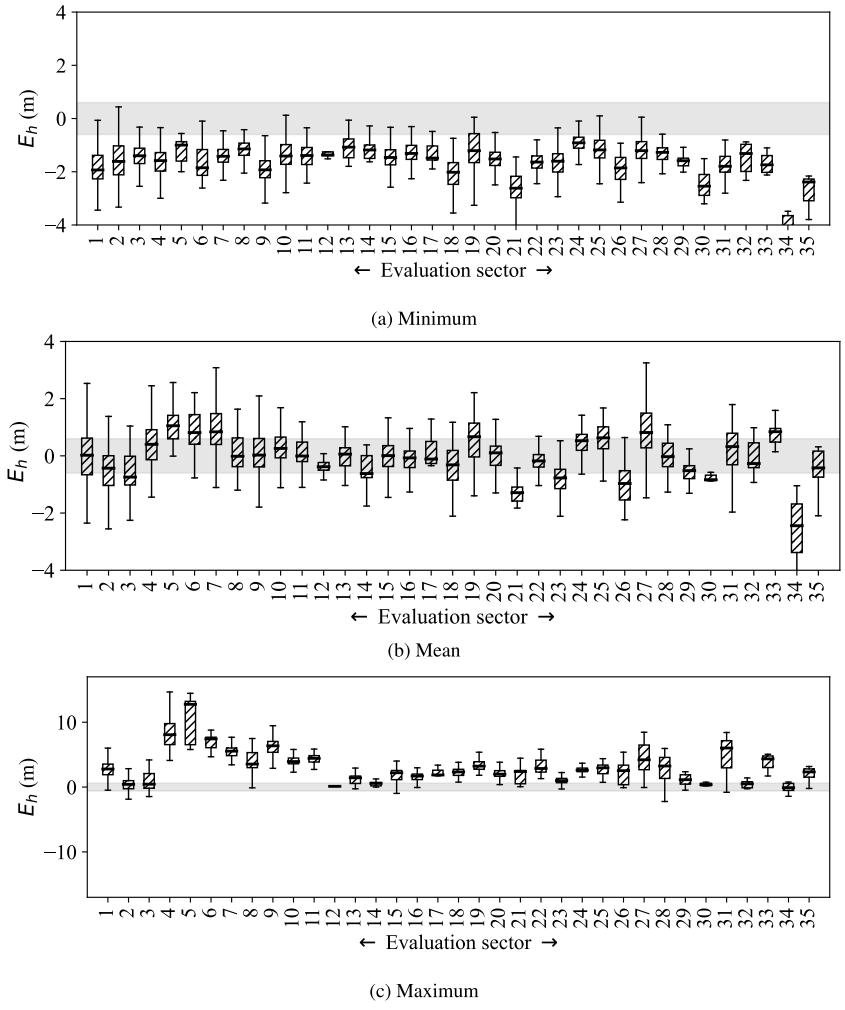
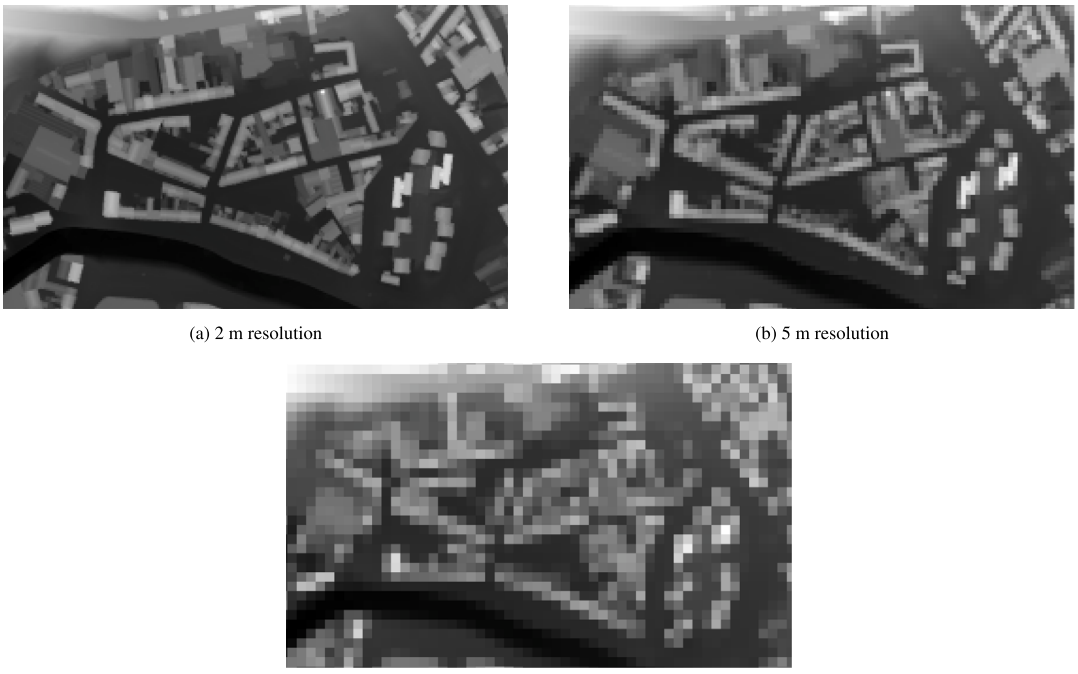


Fig. A.10. Box and whisker plot of WSE difference (point-based) for different resampling operators and 10 m grid resolution. Note: Vertical scale is altered for sub-figure (c).

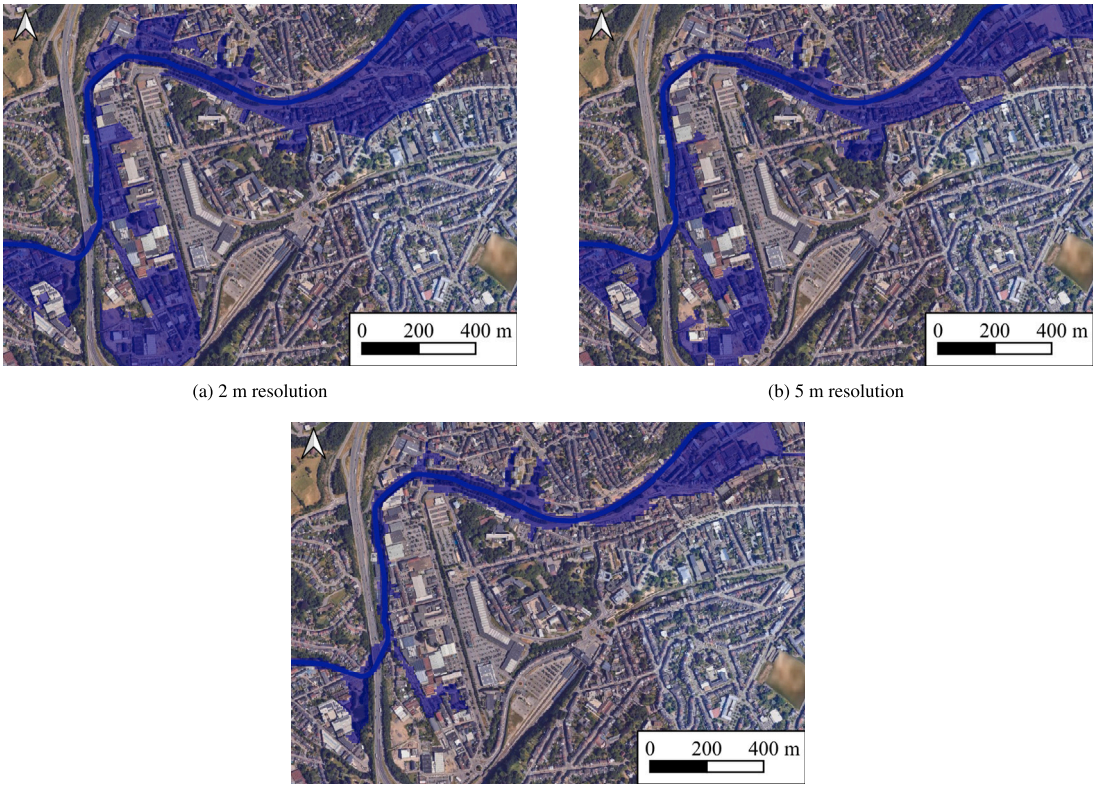


(a) 2 m resolution

(b) 5 m resolution

(c) 10 m resolution

Fig. A.11. Resolved street networks at varying grid resolutions in sector 8.

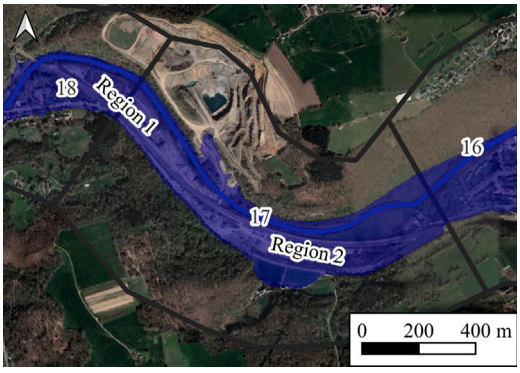


(a) 2 m resolution

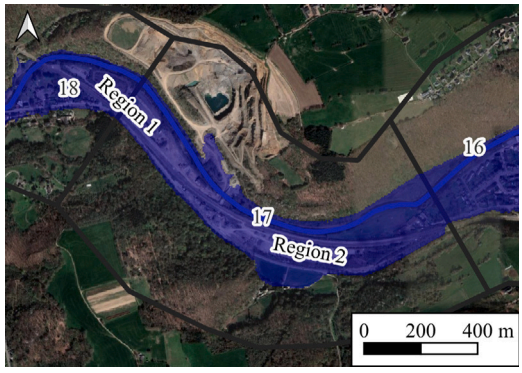
(b) 5 m resolution

(c) 10 m resolution

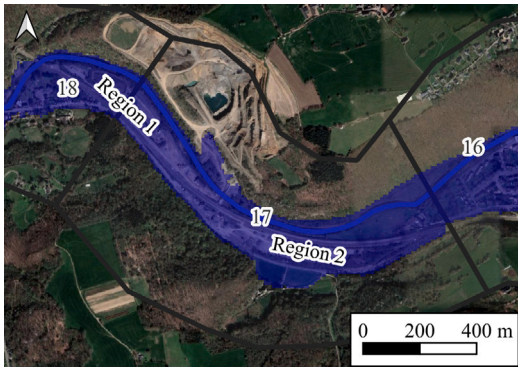
Fig. A.12. FoH method applied to inundation maps of a highly urbanised area (evaluation sector 8) - based on (a) 2 m, (b) 5 m and, (c) 10 m resolution models.



(a) 2 m resolution



(b) 5 m resolution



(c) 10 m resolution

Fig. A.13. FoH method applied to inundation maps of a non-urbanised area (sector 17) - based on (a) 2 m, (b) 5 m and, (c) 10 m resolution models. Note: Black lines indicate evaluation sector boundaries.

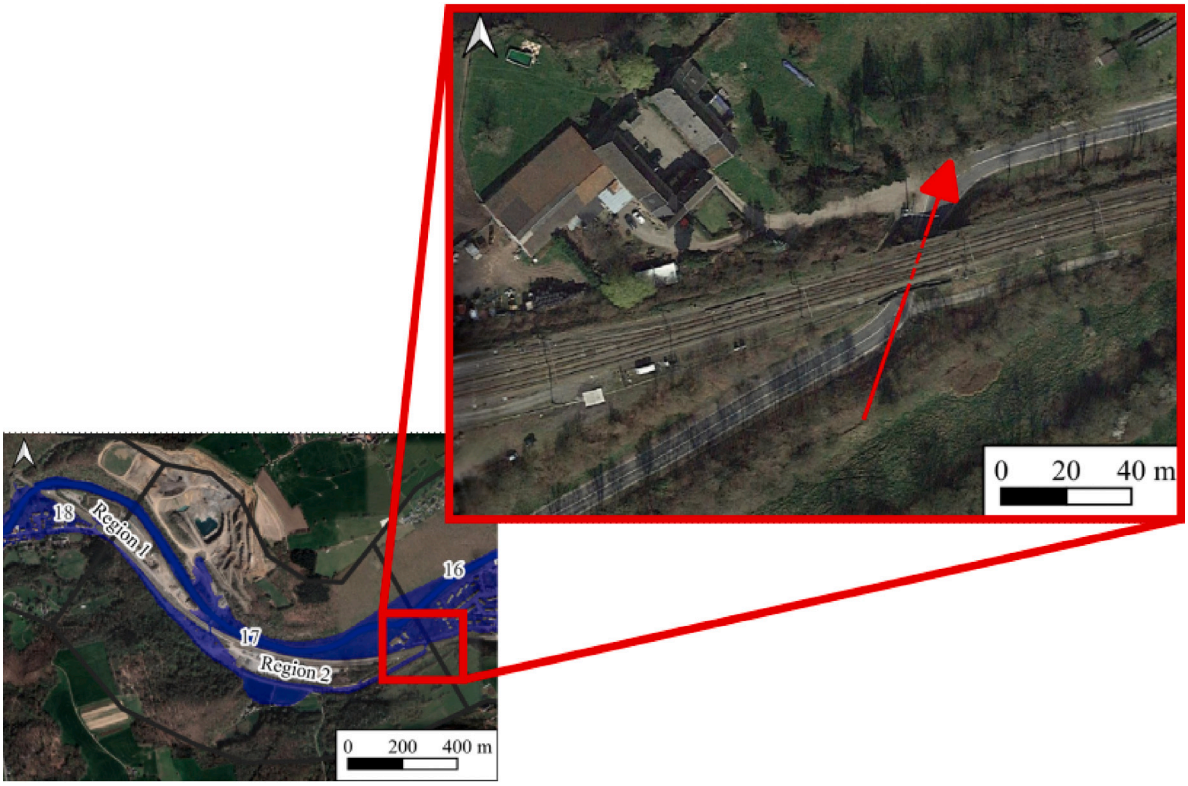


Fig. A.14. Location of railway underpass that was manually introduced in the DEM (sector 17).

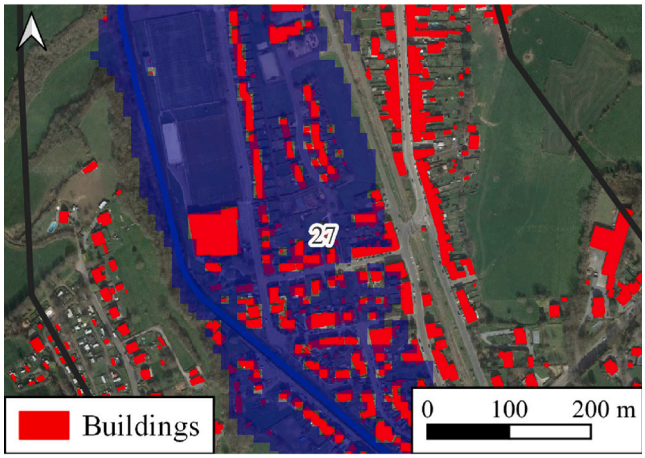


Fig. A.15. IA method applied to the observed inundation map (sector 27).

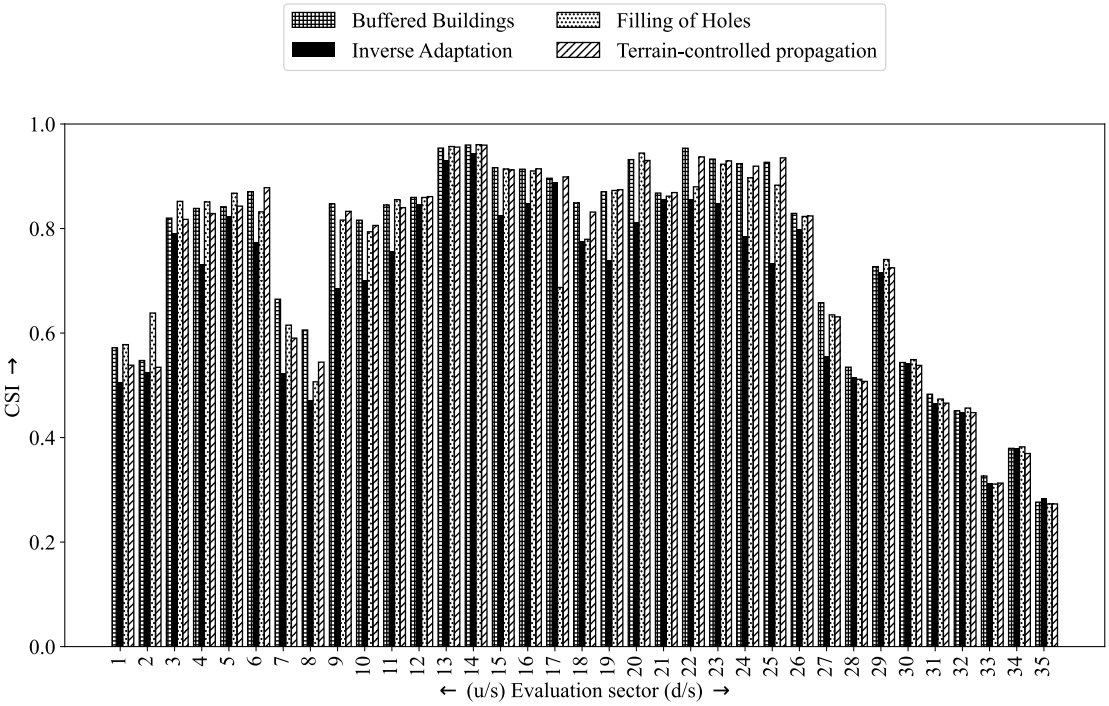


Fig. A.16. CSI for evaluation sectors for case 10m-min.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jhydrol.2025.134578.

Data availability

Data will be made available on request.

References

Alarcón, B., Saenger, V., Guerra, M., Faúndez, R., Link, F., Carrasco, J.A., Link, O., 2024. Using social cartographies for the calibration of two-dimensional hydraulic flood models. Nat. Hazards http://dx.doi.org/10.1007/s11069-024-06838-0.

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Wyser, K., Feyen, L., 2017. Global projections of river flood risk in a warmer world. Earth's Futur. 5 (2), 171–182. http://dx.doi.org/10.1002/2016EF000485.

Alipour, A., Jafarzadegan, K., Moradkhani, H., 2022. Global sensitivity analysis in hydrodynamic modeling and flood inundation mapping. Environ. Model. Softw. 152, http://dx.doi.org/10.1016/j.envsoft.2022.105398.

Anon, 2021. L'information géographique et la géomatique au service des inondations de juillet 2021 | Géoportail de la Wallonie. URL: https://geoportail.wallonie.be/home/ressources/autour-du-geoportail/linformation-geographique-et-la-geomatique-au-service-des-inondations-de-juillet-2021.html.

Archambeau, P., Dewals, B., Erpicum, S., Pirotton, M., Teller, J., Schellings, C., 2022a. Zone d'aléas et périodes de retour. In Schéma stratégique multidisciplinaire du bassin versant de la Vesdre : Diagnostic. Technical Report, Studio Paola Viganò and Team Vesdre-ULiège, pp. 45–46, URL: https://territoire.wallonie.be/storage/territoire/documents/content/page/inondations/schema-vesdre/cartes-et-rapports/vesdre-rapport-diagnostic.pdf.

Archambeau, P., Erpicum, S., Dewals, B., Pirotton, M., Schellings, C., 2022b. Zone d'aléas et périodes de retour. In Schéma stratégique multidisciplinaire du bassin versant de la Vesdre : Diagnostic. Technical Report, Studio Paola Viganò and Team Vesdre-ULiège, pp. 96–113.

Bates, P., 2023. Fundamental limits to flood inundation modelling. Nat. Water 1 (7), 566–567. http://dx.doi.org/10.1038/s44221-023-00106-4.

Beckers, A., Dewals, B., Erpicum, S., Dujardin, S., Detrembleur, S., Teller, J., Pirotton, M., Archambeau, P., 2013. Contribution of land use changes to future flood damage along the river Meuse in the Walloon region. Nat. Hazards Earth Syst. Sci. 13 (9), 2301–2318. http://dx.doi.org/10.5194/nhess-13-2301-2013.

Begnudelli, L., Sanders, B.F., 2007. Conservative wetting and drying methodology for quadrilateral grid finite-volume models. J. Hydraul. Eng. 133 (3), 312–322. http://dx.doi.org/10.1061/(ASCE)0733-9429(2007)133:3(312).

Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Model. Softw. 40, 1–20. http://dx.doi.org/10.1016/j.envsoft.2012.09.011.

Bertola, M., Blöschl, G., Bohac, M., Borga, M., Castellarin, A., Chirico, G.B., Claps, P., Dallan, E., Danilovich, I., Ganora, D., Gorbachova, L., Ledvinka, O., Mavrova-Guirguinova, M., Montanari, A., Ovcharuk, V., Viglione, A., Volpi, E., Arheimer, B., Aronica, G.T., Bonacci, O., Čanjevac, I., Csik, A., Frolova, N., Gnandt, B., Gribovszki, Z., Gül, A., Günther, K., Guse, B., Hannaford, J., Harrigan, S., Kireeva, M., Kohnová, S., Komma, J., Kriauciuniene, J., Kronvang, B., Lawrence, D., Lüdtke, S., Mediero, L., Merz, B., Molnar, P., Murphy, C., Oskoruš, D., Osuch, M., Parajka, J., Pfister, L., Radevski, I., Sauquet, E., Schröter, K., Šraj, M., Szolgay, J., Turner, S., Valent, P., Veijalainen, N., Ward, P.J., Willems, P., Zivkovic, N., 2023. Megafloods in europe can be anticipated from observations in hydrologically similar catchments. Nat. Geosci. 16 (11), 982–988. http://dx.doi.org/10.1038/s41561-023-01300-5.

Bhola, P.K., Nair, B.B., Leandro, J., Rao, S.N., Disse, M., 2019. Flood inundation forecasts using validation data generated with the assistance of computer vision. J. Hydroinformatics 21 (2), 240–256. http://dx.doi.org/10.2166/hydro.2018.044.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R.A.P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G.T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G.B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T.R., Kohnová, S., Koskela, J.J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J.L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., Živković, N., 2019. Changing climate both increases and decreases European river floods. Nature 573 (7772), 108–111. http://dx.doi.org/10.1038/s41586-019-1495-6.

Casulli, V., 2009. A high-resolution wetting and drying algorithm for free-surface hydrodynamics. Internat. J. Numer. Methods Fluids 60 (4), 391–408. http://dx.doi.org/10.1002/fld.1896.

Casulli, V., Stelling, G.S., 2011. Semi-implicit subgrid modelling of three-dimensional free-surface flows. Internat. J. Numer. Methods Fluids 67 (4), 441–449. http://dx.doi.org/10.1002/fld.2361.

Chakraborty, P., De Kock, S., Archambeau, P., Pirotton, M., Erpicum, S., Dewals, B., 2024. Assessing the impact of a large multi-purpose reservoir on flood control under moderate and extreme flood conditions. Hydrol. Sci. J. http://dx.doi.org/10.1080/02626667.2024.2420007.

de Arruda Gomes, M.M., de Melo Verçosa, L.F., Cirilo, J.A., 2021. Hydrologic models coupled with 2D hydrodynamic model for high-resolution urban flood simulation. Nat. Hazards 108 (3), 3121–3157. http://dx.doi.org/10.1007/s11069-021-04817-3.

de Moel, H., Jongman, B., Kreibich, H., Merz, B., Penning-Rowsell, E., Ward, P.J., 2015. Flood risk assessments at different spatial scales. Mitig. Adapt. Strat. Glob. Chang. 20 (6), 865–890. http://dx.doi.org/10.1007/s11027-015-9654-z.

Dessers, C., Archambeau, P., Erpicum, S., Dewals, B., Pirotton, M., 2025. Hydrological modelling of the 2021 mega-flood in the east of Belgium. http://dx.doi.org/10.2139/ssrn.5266922, Available At SSRN.

Dewals, B., Erpicum, S., Pirotton, M., Archambeau, P., 2021. July 2021 extreme floods in the Belgian part of the Meuse basin. Hydrolink 4, 104–107, URL: https://www.iahr.org/library/hydrolink?hid=412.

Ernst, J., Dewals, B.J., Detrembleur, S., Archambeau, P., Erpicum, S., Pirotton, M., 2010. Micro-scale flood risk analysis based on detailed 2D hydraulic modelling and high resolution geographic data. Nat. Hazards 55 (2), 181–209. http://dx.doi.org/10.1007/s11069-010-9520-y.

Erpicum, S., Dewals, B., Archambeau, P., Pirotton, M., 2010. Dam break flow computation based on an efficient flux vector splitting. J. Comput. Appl. Math. 234 (7), 2143–2151. http://dx.doi.org/10.1016/j.cam.2009.08.110.

Erpicum, S., Poppema, D., Burghardt, L., Benet, L., Wüthrich, D., Klopries, E.M., Dewals, B., 2024. A dataset of floating debris accumulation at bridges after july 2021 flood in Germany and Belgium. Sci. Data 11 (1), 1092. http://dx.doi.org/10.1038/s41597-024-03907-8.

Falter, D., Vorogushyn, S., Lhomme, J., Apel, H., Gouldby, B., Merz, B., 2013. Hydraulic model evaluation for large-scale flood risk assessments. Hydrol. Process. 27 (9), 1331–1340. http://dx.doi.org/10.1002/hyp.9553.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S., 2013. Global flood risk under climate change. Nat. Clim. Chang. 3 (9), 816–821. http://dx.doi.org/10.1038/nclimate1911.

Jacob, X.K., Bisht, D.S., Chatterjee, C., Raghuwanshi, N.S., 2020. Hydrodynamic modeling for flood hazard assessment in a Data Scarce Region: a case study of Bharathapuzha River Basin. Environ. Model. Assess. 25 (1), 97–114. http://dx.doi.org/10.1007/s10666-019-09664-y.

Journée, M., Goudenhoofdt, E., Vannitsem, S., Delobbe, L., 2023. Quantitative rainfall analysis of the 2021 mid-july flood event in Belgium. Hydrol. Earth Syst. Sci. 27 (17), 3169–3189. http://dx.doi.org/10.5194/hess-27-3169-2023, https://hess.copernicus.org/articles/27/3169/2023/, https://hess.copernicus.org/articles/27/3169/2023/hess-27-3169-2023.pdf.

Kastridis, A., Kirkenidis, C., Sapountzis, M., 2020. An integrated approach of flash flood analysis in ungauged mediterranean watersheds using post-flood surveys and unmanned aerial vehicles. Hydrol. Process. 34 (25), 4920–4939. http://dx.doi.org/10.1002/hyp.13913.

Khosh Bin Ghomash, S., Apel, H., Caviedes-Voullième, D., 2024. Are 2D shallow-water solvers fast enough for early flood warning? A comparative assessment on the 2021 Ahr valley flood event. Nat. Hazards Earth Syst. Sci. 24 (8), 2857–2874. http://dx.doi.org/10.5194/NHESS-24-2857-2024.

Khosh Bin Ghomash, S., Devi Nallasamy, N., Apel, H., 2025. How can building representation influence flood hazard and impact modelling: Insights from the 2021 Ahr Valley Flood. EGUsphere 2025, 1–25. http://dx.doi.org/10.5194/egusphere-2025-2304, https://egusphere.copernicus.org/preprints/2025/egusphere-2025-2304/, https://egusphere.copernicus.org/preprints/2025/egusphere-2025-2304/egusphere-2025-2304.pdf.

Koenig, T.A., Bruce, J.L., O'Connor, J., McGee, B.D., Holmes Jr., R.R., Hollins, R., Forbes, B.T., Kohn, M.S., Schellekens, M., Martin, Z.W., Peppler, M.C., 2016. Identifying and preserving high-water mark data. Technical Report, U S Geological Survey, Reston, VA, p. 60. http://dx.doi.org/10.3133/tm3A24, URL: https://pubs.usgs.gov/publication/tm3A24.

Liang, Q., Marche, F., 2009. Numerical resolution of well-balanced shallow water equations with complex source terms. Adv. Water Resour. 32 (6), 873–884. http://dx.doi.org/10.1016/j.advwatres.2009.02.010.

Ludwig, P., Ehmele, F., Franca, M.J., Mohr, S., Caldas-Alvarez, A., Daniell, J.E., Ehret, U., Feldmann, H., Hundhausen, M., Knippertz, P., Küpfer, K., Kunz, M., Mühr, B., Pinto, J.G., Quinting, J., Schäfer, A.M., Seidel, F., Wisotzky, C., 2023. A multi-disciplinary analysis of the exceptional flood event of july 2021 in central Europe - Part 2: Historical context and relation to climate change. Nat. Hazards Earth Syst. Sci. 23 (4), 1287–1311. http://dx.doi.org/10.5194/nhess-23-1287-2023.

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J.C.J.H., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U., Macdonald, E., 2021. Causes, impacts and patterns of disastrous river floods. Nat. Rev. Earth Environ. 2 (9), 592–609. http://dx.doi.org/10.1038/s43017-021-00195-3.

Mignot, E., Paquier, A., Haider, S., 2006. Modeling floods in a dense urban area using 2D shallow water equations. J. Hydrol. 327 (1–2), 186–199. http://dx.doi.org/10.1016/j.jhydrol.2005.11.026.

Molinari, D., De Bruijn, K.M., Castillo-Rodríguez, J.T., Aronica, G.T., Bouwer, L.M., 2019. Validation of flood risk models: Current practice and possible improvements. Int. J. Disaster Risk Reduct. 33, 441–448. http://dx.doi.org/10.1016/j.ijdrr.2018.10.022.

Mosimann, M., Kauzlaric, M., Schick, S., Martius, O., Zischg, A.P., 2024. Evaluation of surrogate flood models for the use in impact-based flood warning systems at national scale. Environ. Model. Softw. 173, http://dx.doi.org/10.1016/j.envsoft.2023.105936.

Mustafa, A., Bruwier, M., Archambeau, P., Erpicum, S., Pirotton, M., Dewals, B., Teller, J., 2018. Effects of spatial planning on future flood risks in urban environments. J. Environ. Manag. 225, 193–204. http://dx.doi.org/10.1016/j.jenvman.2018.07.090.

Muthusamy, M., Casado, M.R., Butler, D., Leinster, P., 2021. Understanding the effects of digital elevation model resolution in urban fluvial flood modelling. J. Hydrol. 596, http://dx.doi.org/10.1016/j.jhydrol.2021.126088.

Rodríguez Castro, D., Cools, M., Roucour, S., Archambeau, P., Molinari, D., Scorzini, A.R., Dessers, C., Erpicum, S., Pirotton, M., Teller, J., Dewals, B., 2025. Can macro- or meso-scale coping capacity variables improve the classification of building flood losses? Nat. Hazards http://dx.doi.org/10.1007/s11069-025-07123-4, URL: https://link.springer.com/10.1007/s11069-025-07123-4.

Rulot, F., Dewals, B., Erpicum, S., Archambeau, P., Pirotton, M., 2012. Modelling sediment transport over partially non-erodible bottoms. Internat. J. Numer. Methods Fluids 70 (2), 186–199. http://dx.doi.org/10.1002/fld.2684.

Sanders, B.F., Wing, O.E.J., Bates, P.D., 2024. Flooding is not like filling a bath. Earth's Futur. 12 (12), e2024EF005164. http://dx.doi.org/10.1029/2024EF005164, URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024EF005164.

Savage, J.T.S., Bates, P., Freer, J., Neal, J., Aronica, G., 2016. When does spatial resolution become spurious in probabilistic flood inundation predictions? Hydrol. Process. 30 (13), 2014–2032. http://dx.doi.org/10.1002/hyp.10749.

Scarpino, S., Albano, R., Cantisani, A., Mancusi, L., Sole, A., Milillo, G., 2018. Multitemporal SAR data and 2D hydrodynamic model flood scenario dynamics assessment. ISPRS Int. J. Geo-Inform. 7 (3), http://dx.doi.org/10.3390/ijgi7030105, URL: https://www.mdpi.com/2220-9964/7/3/105.

Schubert, J.E., Sanders, B.F., 2012. Building treatments for urban flood inundation models and implications for predictive skill and modeling efficiency. Adv. Water Resour. 41, 49–64. http://dx.doi.org/10.1016/j.advwatres.2012.02.012, URL: https://www.sciencedirect.com/science/article/pii/S0309170812000425.

Segura-Beltrán, F., Sanchis-Ibor, C., Morales-Hernández, M., González-Sanchis, M., Bussi, G., Ortiz, E., 2016. Using post-flood surveys and geomorphologic mapping to evaluate hydrological and hydraulic models: The flash flood of the Girona River (Spain) in 2007. J. Hydrol. 541, 310–329. http://dx.doi.org/10.1016/j.jhydrol.2016.04.039.

Sethian, J.A., 1999. Fast marching methods. SIAM Rev. 41 (2), 199–235. http://dx.doi.org/10.1137/S0036144598347059.

Shen, Y., Jiang, C., 2023. Quantitative assessment of computational efficiency of numerical models for surface flow simulation. J. Hydroinformatics 25 (3), 782–796. http://dx.doi.org/10.2166/hydro.2023.131.

Shen, J., Tan, F., 2020. Effects of DEM resolution and resampling technique on building treatment for urban inundation modeling: a case study for the 2016 flooding of the HUST campus in wuhan. Nat. Hazards 104 (1), 927–957. http://dx.doi.org/10.1007/s11069-020-04198-z.

Shustikova, I., Domeneghetti, A., Neal, J.C., Bates, P., Castellarin, A., 2019. Comparing 2D capabilities of HEC-RAS and LISFLOOD-FP on complex topography. Hydrol. Sci. J. 64 (14), 1769–1782. http://dx.doi.org/10.1080/02626667.2019.1671982.

Smith, M., Patrick, N., Frazier, N., Kim, J., 2021. Validation of urban flood inundation models applied using nationally available data sets: Novel analyses of observed high water information. J. Hydrol. Eng. 26 (12), http://dx.doi.org/10.1061/(asce)he.1943-5584.0002129.

Tradowsky, J.S., Philip, S.Y., Kreienkamp, F., Kew, S.F., Lorenz, P., Arrighi, J., Bettmann, T., Caluwaerts, S., Chan, S.C., De Cruz, L., de Vries, H., Demuth, N., Ferrone, A., Fischer, E.M., Fowler, H.J., Goergen, K., Heinrich, D., Henrichs, Y., Kaspar, F., Lenderink, G., Nilson, E., Otto, F.E.L., Ragone, F., Seneviratne, S.I., Singh, R.K., Skålevåg, A., Termonia, P., Thalheimer, L., van Aalst, M., Van den Bergh, J., Van de Vyver, H., Vannitsem, S., van Oldenborgh, G.J., Van Schaeybroeck, B., Vautard, R., Vonk, D., Wanders, N., 2023. Attribution of the heavy rainfall events leading to severe flooding in Western Europe during july 2021. Clim. Change 176 (7), 90. http://dx.doi.org/10.1007/s10584-023-03502-7.

Werner, M.G.F., 2004. A comparison of flood extent modelling approaches through constraining uncertainties on gauge data. Hydrol. Earth Syst. Sci. 8 (6), 1141–1152. http://dx.doi.org/10.5194/hess-8-1141-2004, URL: https://hess.copernicus.org/articles/8/1141/2004/.

Wing, O.E., Smith, A.M., Marston, M.L., Porter, J.R., Amodeo, M.F., Sampson, C.C., Bates, P.D., 2021. Simulating historical flood events at the continental scale: Observational validation of a large-scale hydrodynamic model. Nat. Hazards Earth Syst. Sci. 21 (2), 559–575. http://dx.doi.org/10.5194/nhess-21-559-2021.

Yalcin, E., 2020. Assessing the impact of topography and land cover data resolutions on two-dimensional HEC-RAS hydrodynamic model simulations for urban flood hazard analysis. Nat. Hazards 101 (3), 995–1017. http://dx.doi.org/10.1007/s11069-020-03906-z.

Zischg, A.P., Mosimann, M., Bernet, D.B., Röthlisberger, V., 2018. Validation of 2D flood models with insurance claims. J. Hydrol. 557, 350–361. http://dx.doi.org/10.1016/j.jhydrol.2017.12.042.