A note to the editor: One minor mistake was corrected in Figure 1 Panel C, where the position of site B's camera symbol was slightly offset during figure revision. We corrected the position of the camera in the updated manuscript submission and figure file. This change has no bearing on any of the analysis or outcomes of the paper. Additionally, the code availability section, and corresponding reference, were updated to reflect the final published Zenodo repository and DOI. Nothing else was changed relative to the accepted version of the manuscript.

RC1: 'Comment on egusphere-2025-3962', Seyed Mohammad Hassan Erfani, 10 Nov 2025

This is a well-written and methodologically solid paper addressing an important and timely topic—**urban flooding**. The study effectively builds upon previous efforts, particularly Erfani et al. and Eltner et al., and integrates their insights into a novel framework. The authors demonstrate a strong grasp of both the hydrologic and vision-based aspects of flood monitoring, making the work a valuable contribution to the field. Below, I offer a few comments and questions that may help strengthen the manuscript.

We thank the reviewer for the positive assessment of our manuscript. Below, we respond to individual comments, answer questions, and describe updates made to the manuscript to address reviewer comments.

"While aerial lidar offers broad spatial coverage, it does not resolve fine-scale topographic features such as street curbs or shallow depressions common in urban environments" (Dale et al., 2025, p. 7) (pdf)

Why did you use aerial lidar in the first place? If it was not used directly in your workflow, you might consider omitting it to avoid confusion.

[Q1] We thank the reviewer for their question. The aerial lidar serves two key purposes in the analysis:

First, the terrestrial scans are limited in extent, focused on the camera FOV. While their high density is key for water-level extraction, this limits their use for flood extent propagation. As such, the final propagation of a water level to a flood extent is performed on the aerial lidar DTM. The use of a shared elevation dataset for flood-extent propagation enables the evaluation of flood connectivity between the study sites. Additionally, extrapolation beyond a single camera FOV is terrestrial scans at different is necessary in cases where terrestrial scans at different sites do not overlap in extent. We clarify the use of aerial lidar in the methods by adding the following language:

> ***Line 196 - 199*****:** *"Terrestrial lidar scans collected here are confined to areas around each camera location. In order to propagate observed flood-extent beyond the camera FOVs and evaluate connectivity of floodwaters, we integrate the terrestrial lidar scans with the regional aerial survey for St. Clair County."*

> ***Line 332 - 335:*** *"To estimate flood extent beyond the visible portion of the image, we apply an iterative flood-fill procedure to the 0.5 m-resolution USGS DTM (Wu et al., 2018; Samela et al., 2020). As the extent of the terrestrial scans at sites A and B do not overlap, the use of the aerial lidar DTM is necessary for evaluating camera predictions of flood connectivity across multiple sites."*

Second, georeferencing of the terrestrial lidar is evaluated based on elevation differencing with the aerial lidar DTM. Coregistration against a shared elevation dataset ensures consistently in absolute water level between the two study sites. Systematic offsets between aerial and terrestrial lidar are also used to identify potential georeferencing artifacts, such as an induced slope, which are more likely for the small areal extents of the terrestrial scans. This differencing is currently included as a supplemental figure, and we revised the methods to emphasize this step by adding the following:

> ***Line 208 - 210:*** *"Co-registration with the aerial lidar minimizes systematic bias from introduced slopes in the terrestrial lidar and ensures that absolute water levels across sites are in a shared reference, enabling direct comparison of derived extents (Figure S2)".*

"This approach relies on annotated point prompts that indicate the presence or absence of 230 flooding at individual pixels within a reference image." (Dale et al., 2025, p. 9) (pdf)

"For a given flood event, the earliest image in which flooding was visible was annotated with three to five positive point prompts. These prompts were then used to segment the remaining image sequence." (Dale et al., 2025, p. 9) (pdf)

"The visual confirmation of flooding was used to iteratively refine the segmentation, with additional positive prompts added to correct for false negatives (i.e., flooded areas classified as non-flooded), and negative prompts added to address false positives (i.e., non-flooded areas 235 misclassified as flooded)" (Dale et al., 2025, p. 9) (pdf)

I understand that machine learning is not the main focus of this study—it primarily serves as a tool to extract information from 2D imagery. However, given that previous studies have already addressed similar challenges, it might have been advantageous to employ some of those established methods directly. Although the amount of manual annotation here is reduced, it still represents a bottleneck to achieving full automation.

[Q2] We agree with the reviewer that automated flood segmentation with a purpose trained model is vital to practical deployment of camera-based systems. However, we emphasize that the the aim of this study is not a ready-made analysis package, but a more generalized workflow.

For the scope of this contribution, point- prompted SAM is an expedient method to obtain flood masks for the limited number of case study images we present here, but we agree that for operation at scale, a fully automated method could be employed. We have added further emphasis to the discussion that the photogrammetry methods are agnostic to the source of flood masks, and the potential for future integration with fully automated segmentation models by updating the text to reflect the following:

> ***Lines 532-339:*** *"A major advantage of our workflow lies in the modularity of our processing pipeline and the relative ease of camera deployment. SegmentAnything, facilitated the efficient generation of accurate flood masks for our case study images without dedicated model training. Consistent with prior work (Moghimi et al., 2024, Wang et al., 2024), we found that SegmentAnything performed robustly for floodwater segmentation, producing masks with mean IoU>90% in most cases. However, the scalability of manual prompting for long-term automated monitoring is limited (Zamboni et al., 2025). Our new analysis workflow is agnostic to the specific source of flood masks, and future applications could easily incorporate improved segmentation techniques, either domain-specific or fine-tuned foundation models (e.g., Wagner et al., 2023), without changing the overall processing pipeline".*

"The extrinsic camera pose matrix, P, was estimated based on a set of matched reference features with known locations in both image coordinates (u, v), and world coordinates (X, Y, Z). This process, known as the Perspective-n-Point (PnP) problem, yields an estimated camera pose denoted as PPnP. Feature matching was performed manually, with image coordinates of reference features labeled in ImageJ (Schindelin et al., 2012) and their corresponding world coordinates annotated from the terrestrial lidar point cloud using CloudCompare (CloudCompare, 2023). In the absence of permanent ground control points, 270 static scene elements such as rooftops, fence posts, and utility poles were used as reference features. Between 20 and 30 such features were labeled for each camera. Point precision was limited by image resolution, point cloud noise, and the spatial resolution of the lidar scan." (Dale et al., 2025, p. 10) (pdf)

In this section, the methodology appears somewhat behind the state of the art. As mentioned earlier, even though these technical components might seem peripheral, exploring ways to automate them is crucial for advancing toward **operational applications** of such frameworks.

[Q3] We agree that automated image registration approaches, such as GCP matching, have significant benefits for operational deployment (Erfani et al. 2023; Eltner et al. 2021; Jeon et al. 2022). However, part of the novelty of this study is its application in a setting where these constraints are not possible. In our case, practical limitations of installation on resident properties, with public roads made permanent GCP installation impossible. While this does come at the expense of automating aspects of the analysis pipeline, it also emulates the likely constraints of using existing camera infrastructure for flood monitoring, which is a major area of opportunity for expanding vision-based flood monitoring (Wang et al. 2024). We added the following to the text to more clearly articulate this limitation but also highlight the future opportunities in this area:

> ***Lines 556-565: "****Our setup mirrors the likely constraints for working with images from existing public sensors, such as security cameras, that lack explicit ground control. Although the use of static scene features as GCPs can limit labeling precision and the semantic depth of extracted features, it enables scalable, repeatable deployment across heterogeneous camera networks and allows for quantitative flood observations from imagery that would otherwise be unusable for absolute measurements. In contexts requiring higher precision, the use of continuously visible, permanent ground control points (Erfani et al., 2023), fixed-mount cameras (Wang et al., 2024), or onboard inertial measurement units (IMUs) could reduce these uncertainties and allow for the reliable implementation of automated approaches. Future implementations could also integrate automated drift correction or recent advances in machine-learning-based image-to-point cloud registration (Bai et al., 2024; Jeon and Seo, 2022)."*

Also, how many times did the authors perform this procedure? Assuming the camera locations are fixed, it seems unnecessary to repeat it multiple times—unless the cameras were moved between events.

In the case of the two case study events, camera location is static, however camera angle, principally bearing, did drift between events (estimated at 13 degrees for Camera A and 24 degrees for Camera B). As a result, pose estimation was repeated for each event. We will add this detail t the methods section as:

> ***Line 302 - 303:*** *"Camera post positions were stable during the case study events, however, between events camera bearing shifted by approximately 13° at Site A, and 24° at site B, requiring separate pose estimations for each event."*

"A separate camera pose estimate was computed for each camera and flood event. For the moderate May 14 flood, Camera A's pose was calculated using 18 reference features, yielding a median reprojection error of 6.83 pixels. The recovered camera location was offset 46 cm from the labeled camera center in the point cloud. For the July 4 event, pose estimation at Camera A used 24 features, resulting in a median reprojection error of 23.6 pixels and a reduced camera position offset to 6 cm." (Dale et al., 2025, p. 11) (pdf)

This part is a bit confusing. Could the authors clarify why the July event—with more reference features—has a higher reprojection error in image space but a smaller offset in 3D space? The 3D error seems quite large and could significantly affect flood mapping accuracy (e.g., introducing nearly a meter of uncertainty in flood extent). Did the authors examine how this uncertainty propagates into flood depth estimates?

[Q4] We agree with the reviewer that when camera pose estimation relies on reference features that are distant from the camera and span a limited range of depth, ambiguity can arise in the recovered camera position. In such cases, reprojection error in pixel space becomes relatively insensitive to camera translation along the optical axis, allowing multiple candidate poses with similar reprojection error but different 3D camera locations. In such cases, the solver can minimize reprojection error effectively by shifting the camera along the optical axis, resulting in a larger 3D positional offset (46 cm).

To directly assess the impact of this ambiguity on water level estimation, we re-estimated the May 14 camera pose while constraining the camera position to less than 10 cm from the surveyed camera center. This constraint increased the median reprojection error only marginally (from 6.83 to 7.11 pixels) and resulted in a mean difference in estimated water surface elevation of 1.5 cm. This indicates that, while ambiguity in camera translation can affect recovered camera position, its influence on water level estimates is limited for this case, particularly at distances relevant to the extracted flood edges. In contrast, the July event utilized more features (24) with better spatial distribution. This geometric diversity constrained the 3D camera position more

effectively (6 cm offset), even though individual features had higher re-projection error. We have updated the methods to clarify this point:

> ***Line 303-308:*** *"For the moderate May 14 flood, Camera A's pose was calculated using 18 reference features, yielding a median reprojection error of 6.83 pixels. The recovered camera location was offset 46 cm from the labelled camera center in the point cloud. This apparent mismatch reflects the reduced sensitivity of reprojection error to camera position when reference features are distant or near planar, which weakens 3D positional constraints despite low pixel error. Constraining the shift in the May 14 camera position to ≤10 cm increased median reprojection errors only slightly (to 7.11 pixels)."*

"Flood extent estimation is based on the intersection of lidar-derived topography and image-derived water classifications. Using the established projection pipeline in Equation 2, each point in the terrestrial lidar point cloud is mapped to a corresponding image pixel. If a pixel is identified as flooded in the SAM2-derived binary segmentation mask, the associated terrestrial lidar point is classified as inundated." (Dale et al., 2025, p. 11) (pdf)

How was this implemented? Since multiple 3D points may project onto a single image pixel, how did the authors handle indexing or correspondence between flooded pixels and their associated 3D points?

"To estimate water surface elevation (WSE), the highest elevations along the boundary of the inundated zone are used as a proxy for the maximum water level and the water surface is 305 assumed to be flat. Edge pixels are extracted using a Canny Edge Detection filter, and the 90th and 95th percentiles of the extracted edge elevation distribution are used to represent a range of possible water surface levels (WSE90 and WSE95) to account for potential topographic noise or obstruction of the water edge in the time lapse images." (Dale et al., 2025, p. 11) (pdf)

This appears to be the **core contribution** of the paper and would benefit from more detailed elaboration. The rest of the workflow closely follows prior studies.

[Q6] We thank the reviewer for their comment. The reviewer is correct that projecting a dense 3D point cloud into image space can result in non-unique pixel–point mappings, with multiple 3D points projecting onto a single image pixel. Our implementation explicitly accommodates this ambiguity and avoids one-to-one pixel-to-point correspondence. In our approach, all terrestrial lidar points that project to image pixels classified as flooded in the SAM2-derived segmentation mask are retained and labeled as inundated, while preserving their original 3D indices and UTM coordinates. Rather than assigning a single elevation value to each flooded pixel, the full set of inundated 3D points is interpolated into a 0.05 m resolution raster representing the spatial flood extent in map space.

Water surface elevation (WSE) is then estimated from this rasterized flood boundary, not from individual image pixels or raw point projections. Canny edge detection is applied to the

rasterized inundation extent to identify the flood boundary, and the 90th and 95th percentiles of the resulting edge elevation distribution (WSE_{90} and WSE_{95}) are used to estimate water level while accounting for topographic noise and partial occlusion of the water edge in time-lapse imagery. This raster-based formulation reduces bias associated with distance-dependent point density and avoids over-weighting regions where many 3D points project to a single pixel. We have clarified this implementation in the Methods section and revised Figure 3 to explicitly indicate that water level estimation is raster-cell–based rather than point- or pixel-based. The modified text will read:

> ***Line 315-331:*** *"Flood extent estimation is based on the intersection of lidar-derived topography and image-derived water classifications. Using the established projection pipeline in Equation 2, each point in the terrestrial lidar point cloud is mapped to a corresponding image pixel. If the image pixel is identified as flooded in the SAM2-derived binary segmentation mask, the associated terrestrial lidar point is classified as inundated. Because pixel resolution decreases with distance from the camera, multiple 3D points may project onto a single image pixel; therefore, all inundated points are retained and a one-to-one pixel–point correspondence is not enforced. Together, these inundated points represent the portion of the ground surface that is underwater at the time the image was captured. This set of inundated points is interpolated into a 0.05-meter resolution raster representing the visible flood extent in the image. This interpolation step reduces bias associated with distance-dependent differences in point density and avoids over-representation of regions where many 3D points project to a single pixel. Water surface elevation (WSE) is estimated from the rasterized flood boundary rather than from individual pixels or raw point projections. Canny edge detection is applied to the rasterized inundation extent to identify the flood boundary, and the 90th and 95th percentiles of the resulting edge elevation distribution (WSE_{90} and WSE_{95}) to account for potential topographic noise or obstruction of the water edge in the time lapse images. Assuming a flat-water surface, elevations along the flood boundary should exhibit a sharp peak at the upper end of the elevation distribution. The consistency and sharpness of this peak are another parameter useful to evaluate the camera pose estimation, as errors in estimated camera orientation or translation produce unrealistically large elevation differences between near- and far-field water edges."*

Based on Figure 1, I initially thought the authors were using a **hypsometric curve** approach (Dale et al., 2025, p. 6) (pdf). It might be helpful to elaborate on how these curves are utilized and how they relate to the conceptual model applied later in the **iterative flood-fill procedure** at 0.5 m resolution (Wu et al., 2018; Samela et al., 2020).

[Q7] We are sorry for the confusion. The hypsometric curves in Figure 1 are strictly illustrative of broad elevation differences between sites and do not directly factor into the water-level estimation. We will revise the figure to remove the hypsometric curve to avoid confusing the readers. Conceptually the flood-fill procedure will follow the area trajectory shown in these curves with the key distinction being that the hypsometric curves do not consider water level dependent connectivity produced by local depressions. This is the primary advantage of the iterative flood-fill procedure over a simple elevation threshold applied over the entire domain, which will over-predict flood extent for disconnected low elevation areas. We will clarify the advantage of the iterative flood-fill procedure in the methods as follows:

> ***Line 335 –338:*** *"Beginning at the lowest observed elevation within the camera's field of view, adjacent terrain cells are iteratively inundated if their elevation is below the target WSE, continuing until no*

additional cells meet this condition. ***The requirement of topographic connectivity with the seed point prevents over prediction of flood extents likely with a simple elevation threshold applied to the entire domain."***

"The area of interest for the flood-fill implementation focused on the direct area spanning the two camera locations, approximately 500 m by 250 m, to avoid propagation into unobservable areas." (Dale et al., 2025, p. 11) (pdf)

This aspect could also be an interesting avenue for future research—for example, using a **location-allocation optimization** approach to minimize the number of cameras while maximizing the coverage area.

[Q8] We completely agree with the reviewer's comment that network optimization is an important consideration for operational camera monitoring. While beyond the scope of the current study, we will add discussion of key considerations for camera integration based on our interpretation of flood connectivity in the current case-study:

Lines 730 - 733: "*Future research could leverage this framework to optimize camera network configuration, balancing the number and placement of ground-based cameras to maximize spatial coverage and the ability to observe flood connectivity (Negri et al., 2025; Zhao et al., 2025)."*

"Although image data informed general model development, no direct calibration against the imagery was performed." (Dale et al., 2025, p. 12) (pdf)

This raises an interesting question: if sparse information extracted from cameras were available, how could such data be assimilated into flood models to refine their outputs? Could this be implemented in real time?

[Q9] This is an excellent point by the reviewer, and a direction we have been exploring in our current work. We believe specific calibration approaches are beyond the scope of the current paper but will add broader discussion of potential for camera-model integration to the revised discussion as follows:

Line 676-691: "*Camera-based observations provide a promising avenue to address these calibration gaps. Depending on the data available and the precision required, camera-derived information could support multiple levels of model calibration. At a minimum, observations of flood presence, extent, and connectivity can serve as semi-quantitative validation of model structure and behavior. More detailed or well-distributed camera installations could function as stream-gauge surrogates, enabling direct calibration of key model parameters such as surface roughness, stormwater capacity, or flood wave timing. These approaches could ultimately facilitate both post-event model evaluation and real-time model adjustment, bridging gaps in empirical data for urban flood forecasting.*

When possible, to implement, camera-derived WSEs offer a rare empirical reference for validating modeled spatiotemporal patterns of inundation. For example, these high-resolution, time-resolved observations enabled direct comparison with outputs from an uncalibrated HEC-RAS Rain-on-Grid simulation of the July 4 flood event, revealing a close match in peak flood depth, timing, and extent. This

proof-of-concept highlights the strong potential of integrating image-derived data into calibration workflows for 2D hydrodynamic models, particularly in high-flow scenarios where floodwaters are hydraulically connected and drainage networks are overwhelmed. Beyond event reconstruction, such observations can support real-time model updating, performance evaluation of stormwater infrastructure, and planning for flood mitigation in poorly instrumented or rapidly evolving urban settings, providing a practical, data-driven way to reduce uncertainty in urban flood simulations."

"Our comparison focuses on quantifying the relative agreement in predicted flood extent between the two methods. The primary metric focuses on identifying regions where both the model and camera-based approaches indicate flooding areas of mutual agreement in predicted inundation. This shared extent is expressed as Foverlap, the ratio of the number of pixels classified as flooded by both methods to the total number of pixels classified as flooded by either. The model domain includes areas separated from our camera sites by major roads and drainage canals. To provide a meaningful comparison between model output and our image-based methods, we spatially restricted our comparison to a region with the approximate bounds of the topographic depression containing the study neighborhood. Where flood extents overlap, we also compared modeled and observed water surface elevations and flood depths." (Dale et al., 2025, p. 12) (pdf)

This section feels somewhat unconventional and could benefit from clarification. If I were the authors, I would consider treating the HEC-RAS output as the **reference (or ground truth)** and evaluating the vision-based estimates using standard metrics such as a **confusion matrix**. This would make the comparison more transparent and interpretable.

[Q11] We agree that additional metrics of will improve the interpretation of the model-camera comparison. However, we want to be explicit that the model-camera comparison is intended to compare key differences in the behavior assumptions of the two approaches, rather than treating one or the other as an absolute reference for the purposes of validation. We will add clarification of this distinction:

Lines 370-373: "*Because the model itself is only qualitatively calibrated, its output is not treated as a direct validation for absolute water levels estimated from images. Instead, it characterizes similarity or divergence in flood behavior predicted by each method. This is quantified both in terms of the relative agreement in predicted flood extent, and spatial flood connectivity, between the two methods."*

We will add two additional metrics of the asymmetrical agreement between camera and model predictions. These are computationally equivalent to standard precision & recall metrics, but is framed as agnostic to ground truth:

$F_{overlap}$: Symmetric agreement between camera and model (equivalent to IoU)

$F_{M|C}$ *:* Fraction of camera predicted extent predicted by the model

$F_{C|M}$*:* Fraction of model flood extent predicted by the camera

We will add these metrics to the discussion of the model comparison:

Lines:376-378: "*The relative underprediction and/or overprediction of each method is expressed as the fraction of camera predicted flood cells predicted by the model ($F_{M|C}$), and the fraction of model predicted flood cells predicted by the camera ($F_{C|M}$)."*

Line 487-517: "*Benchmarking the hydrodynamic flood model against flood-fill results shows generally good agreement in both the progression and extent of flooding, particularly during the July 4 event. Comparisons metrics $F_{overlap,}$ $F_{M|C,}$ and $F_{C|M}$ were calculated for the peak flood extent for each event. Full confusion matrices of flood area agreement are provided in SI Figure 6. The model successfully captures the broad dynamics of inundation, though key differences emerge in the spatial structure and timing of connectivity between flood patches. During the May 14 event, the model produces disconnected flood patches, even at peak flooding, qualitatively consistent with observations at Site A (Fig 7a). However, the model does not capture the short period of connectivity estimated by flood-fill propagation from both Sites A and B (Figure 7b). Of the modeled May 14 flood extent, eleven patches exceeded 100 m², accounting for 42% of the total modeled inundated area (A_{model}=0.7 ×10⁴ m²), suggesting a bias toward small, isolated flood zones. Agreement metrics for the May 14 event reflect this fragmentation (Figure 7a, b): comparing the peak flood-fill extent based on the Site A WSE_{90}, $F_{overlap}$=0.24, with modeled extent 43% lower than the flood-fill. At Site B the total agreement between the peak flood-fill extent was similar ($F_{overlap}$=0.21), with modeled extent 63% lower in area. The effect of these isolated patches is seen in the flood-fill extent at Site A still only covering 54% ($F_{C|M}$ = 0.54) of the total model extent, but 95% of the largest model flood patch. Similarly at Site B, $F_{C|M}$ = 0.64 for the full extent, and 0.99 for the largest flood patch. A similar effect is seen limiting the comparison to the camera FOV, with $F_{overlap}$ increasing to 0.34 and 0.26, and $F_{C|M}$ increasing to 0.88 and 0.99 for Sites A & B respectively. These spatial mismatches are accompanied by consistent underestimation in water surface elevation, with median modeled values 22 cm and 25 cm below WSE_{90} at Sites A and B, respectively. Depth difference maps highlight these discrepancies, with distinct peaks aligning with the isolated modeled patches (Figure 7c)*

In contrast, the estimated flood extents from the rain-on-grid model and our new method demonstrate significantly closer agreement for the more severe, July 4 event. At the peak of camera observed flooding, the model predicts a single contiguous flood patch, accounting for 80% of the total modeled inundated area (A_{model}=3.86 ×10⁴ m²) and connecting Sites A and B (Figure 7d, e). Flood-fill-model agreement was also significantly higher for the July 4 event with $F_{overlap}$=0.79 and 0.77 for Sites A and B, respectively. Restricting the comparison to the within the camera FOV increases agreement to $F_{overlap}$=0.90 and 0.96 at Sites A and B, respectively. The model prediction near fully encompasses the flood-fill extents from both Site A ($F_{M|C}$ = 0.98), and Site B ($F_{M|C}$ = 1.00), while remaining isolated flood patches in the model limit flood-fill extents to 80 % ($F_{C|M}$ = 0.80) and 77 % ($F_{C|M}$ = 0.77) of model predication at each site. As with the moderate event, limiting the comparison to the largest model flood patch, increases $F_{C|M}$ to 0.93 for Site A, and 0.88 for Site B. Aside from minor edge effects, the model reproduces a nearly flat water surface, with a difference of only 0.5 cm in the mean predicted water surface elevation within the camera FOV at Site A & Site B. The median water surface elevation of elevations of 126.29 m is just 1 cm below the WSE_{90} at Site A and 1 cm above at Site B and yielding closely matched depth distributions (Figure 7f)."

We have additionally added confusion matrices as supplementary figures:

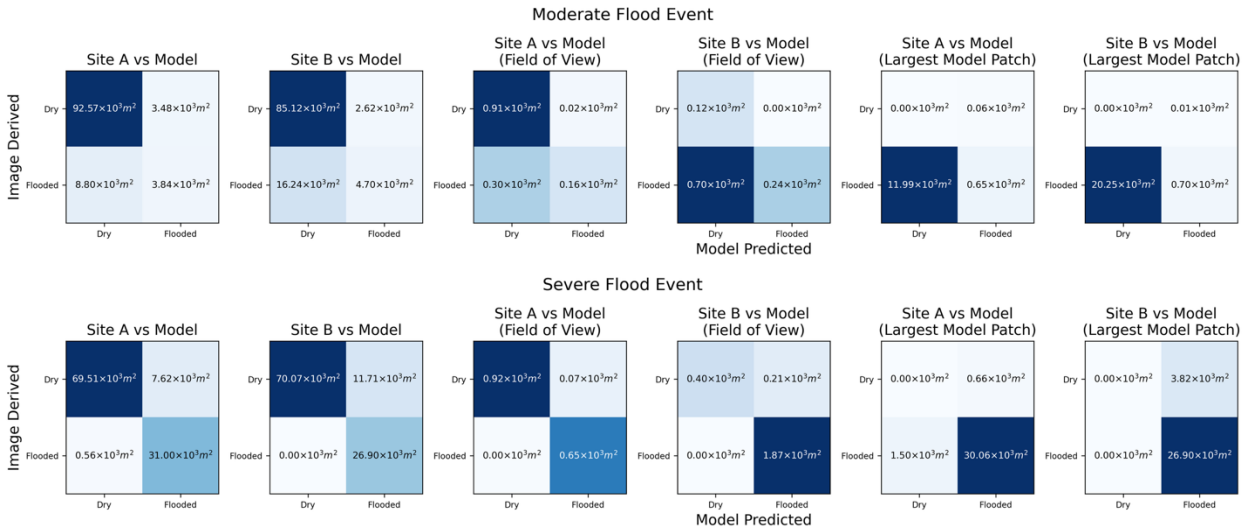

Figure S4: Confusion matrices for image-derived and 2D model predicted peak flood extents. Values indicate the area of overlap and disagreement in flooded and non-flooded raster cells between the methods. Comparisons are shown for the full model extent, restricted to the camera FOV, and restricted to the largest contiguous patch of model predicted flooding.

It would also help highlight that the vision framework is not isolated—the overall performance reflects both the errors of the camera-based method (which provides boundary and initial conditions) and those of the conceptual flood model. A more detailed characterization of each component's contribution would strengthen the paper considerably.

[Q12] We agree with the reviewer that error propagation for individual steps of the analysis pipeline, including image processing and flood-extent estimated is an important element of the usability of the method, and we have added a more explicit treatment to the discussion. We have also added additional notes regarding performance of the image projection to the methods in response to previous reviewer comments. We added additional details uncertainty propagation between the water level estimation and flood extent generation components of our method in Lines 641-660:

> *"Elements of our method – including spatial aggregation of flood boundaries and the calculation of multiple water-level thresholds – effectively constrain uncertainty to levels suitable for urban flood characterization. We find that water-level estimates are robust to both minor random and systematic errors in flood-mask segmentation. For the severe event at Site A, water levels calculated for the half of the scene containing a parked car that partially obscured the water line differed by less than 2 cm, on average, from estimates derived from the unobstructed portion of the scene. Similarly, introducing random jitter of 10–20 pixels to the flood-mask boundary produced mean water-level differences of less than 2 cm. More substantial errors in flood segmentation or camera pose that are not mitigated by the method are typically identifiable through diagnostic artifacts, including large reprojection errors, asymmetric projected flood extents, or exaggerated differences between WSE90 and WSE95. In addition, the visual context provided by the images allows qualitative validation against observable flood indicators such as roadway overtopping.*
>
> *Beyond water-level uncertainty, flood-extent estimates are influenced by the quality of the underlying topographic data. Even with careful georeferencing, physical landscape change between*

surveys or differences in lidar point density can introduce localized elevation discrepancies. Flood extents propagated using aerial lidar tend to be biased toward overprediction because fine-scale topographic structures, such as curbs or drainage ditches, are only partially resolved. Both water-level estimation and flood-extent propagation may therefore be most sensitive at lower water levels, where small-scale topographic features exert stronger control on flood extent. Consequently, interpretations of discrete changes in flood connectivity resulting from small increases in water level should be treated cautiously. However, because the camera images directly capture the spatial distribution of floodwaters between sites, they can provide an independent observational check on modelled flood connectivity and allow clear identification of locations where modelled inundation diverges from observed flooding. Future work should further investigate how uncertainty propagates among these sources."

RC2: 'Comment on egusphere-2025-3962', Anonymous Referee #2, 11 Dec 2025

In their paper, the authors present an innovative urban flood monitoring approach. Intersecting segmented flood masks derived from imagery recorded by low-cost trail cameras and lidar data, they estimate flood water surface elevations for two flood events. Maximum flood depths and extents were then compared with results from a 2D hydrodynamic model.

I have read the paper with interest and think it can be published after major revision. My detailed comments are included below.

Major concerns

1. HEC-RAS model

While I agree in general that comparing flood extent and/or depth from the authors' new method with results from a 2D hydrodynamic model might be an interesting analysis, the paper in its current for lacks important details regarding how the HEC-RAS model was implemented:

[Q1] The details of the HEC-RAS model implementation are currently included in the Zenodo data supplement. Because the HEC-RAS model itself is secondary to the development of our analysis pipeline, we do not want to overwhelm the readers with too many details in the main text given the current length of

the manuscript. We will migrate the relevant elements into the main text methods section and supplementary information.

How was the model grid set up?

We have added the following to the supplement:

> ***(Text S1) Model domain:*** *The computational mesh was generated from a TIN-interpolated, and gap filled, 0.5 m resolution DTM generated from 2019 USGS 3DEP aerial lidar data. The base mesh was generated with 10 m node spacing. Breaklines were added for channel centerlines, culvert inflows and storm drains. Mesh refinement was applied within 5 m of these breaklines, reducing node spacing to 1 m for the area of interest.*

What infiltration method was used in the rain-on-grid approach, and how was it parameterized? What land use classifications and corresponding roughness coefficient were used?

We have added the following to the supplement:

> ***(Text S1) Mannings roughness (n) and runoff:*** *30 m resolution National Land Cover Database (NLCD) were used to define spatially variable roughness coefficients, using HEC-RAS manual reference values for each classification. This was refined using vector polygons of road surfaces and building footprints from the Illinois Department of Transportation. Within building footprints n was assigned a high value of 10 which prevents the routing of runoff from those cells.*

Land Cover Classification	Manning's N Value
NoData	0.035
Roads	0.01
Buildings	10
Banks	0.04
MainChannel	0.03
Open Water	0.035
Developed, Open Space	0.035
Developed, Low Intensity	0.08
Developed, Medium Intensity	0.12
Developed, High Intensity	0.15
Barren Land Rock/Sand/Clay	0.025
Deciduous Forest	0.15
Evergreen Forest	0.15

Mixed Forest	0.1
Shrub/Scrub	0.9
Grassland/Herbaceous	0.04
Pasture/Hay	0.045
Cultivated Crops	0.05
Woody Wetlands	0.07
Emergent Herbaceous Wetlands	0.07

Table S2: Reference landcover based Manning 's roughness coefficients taken from USACE 2024.

(Text S1) Rainfall Runoff: *The curve number (CN) method was used to calculate initial infiltration losses and runoff generation. The same NLCD landcover, and IDOT building and road layers were used to define spatially variable values for CN, abstraction ratio and minimum infiltration rate. Reference values were taken from the HEC-RAS hydraulics manual, with abstraction ratios suggested by Hawkins and Jiang 2023. No additional infiltration losses are calculated after the initial rainfall-runoff conversion.*

Name (Land Cover: Soil Hydric Group)	**Curve Number**	**Initial Abstraction Ratio**	**Abstraction Ratio**	**Minimum Infiltration Rate**
Developed, Medium Intensity : D	86	0.05	0.082	1.270
Developed, Open Space : B/D	74	0.05	0.176	3.485
Emergent Herbaceous Wetlands : A	76	0.05	0.158	7.600
Grassland/Herbaceous : D	80	0.05	0.125	1.270
Buildings :	100	0.05	0.000	0.000
Roads : C	98	0.05	0.010	0.000

Was storm drain infrastructure modeled, or only surface flow?

We have added the following to the supplement:

(Text S1) Stormwater system: *The combined sewer-stormwater system was modeled as a 1D pipe network in HEC-RAS. The locations of stormwater inflows were taken from the Illinois Department of Natural Resources (IDNR) and Heartlands Institute survey of the Prairie Du Pont Watershed. Precise information on the topology and hydraulics of the sewer-stormwater system is unavailable, and connections between inflows were inferred based on published IDNR and USACE reports and maps (USACE 2024; IDNR 2023). Pipe diameter of 0.7 m, based on IDNR survey, and n of 0.015 m were used, based on USACE reported values. The pipe network was connected to the know drainage ditch outfall.*

Are there storage areas within the model domain?

> Because the study area is separated from nearby lakes and reservoirs by the major drainage ditches and roads no additional storage areas were included.

Without additional detail, a review of this portion of the analysis is nearly impossible. Even if detail is added, I still question the value of comparing the authors results with those from an uncalibrated HEC-RAS model; I also suspect the lack precipitation data form within the study area adds substantial uncertainty to model results (it sounds like data from only one rain gauge was used for each event, and gauges were located at a distance of 6 and 8 km from the study area, respectively).

[Q3] We thank the reviewer for their comments, but even without quantitative calibration we believe there is significant value in our model comparison. A major motivation of our study is monitoring approaches for areas without sufficient data for pluvial model calibration. This approach is not unique to our study. There are multiple prior studies where uncalibrated, and otherwise simplified 2D models are used to evaluate new DTM based methods (e.g. Samela et al. 2020; Preisser et al. 2022). To that end, the comparison is intended to identify major similarities and differences in characteristic behavior between camera-derived flood extents and a rain-on-grid model, not as an absolute ground-truth. While factors like precipitation uncertainty will modify maximum flood timing and extent, they are unlikely to alter the behaviors implicit to a rain-on-grid model which distinguish it from our image-based estimates. We will revise the methods and discussion to clarify our conceptual approach and qualify the limits of direct camera to model validation:

> ***Lines 370-373:*** *"Because the model itself is only qualitatively calibrated, its output is not treated as a direct validation for absolute water levels estimated from images. Instead, it characterizes similarity or divergence in flood behavior predicted by each method, based. This is quantified both in terms of the relative agreement in predicted flood extent, and spatial flood connectivity, between the two methods."*

> ***Lines 701-709:*** *"Despite these challenges, our results demonstrate how empirically-derived WSEs can complement and strengthen traditional hydraulic modeling workflows. Our method provides continuous, high-resolution estimates of water level and extent that are directly tied to real flood behavior, capturing sub-decimeter changes in WSE and floodwater connectivity that would otherwise be missed by point-based flood monitoring and modeling approaches. While further validation of camera-derived extents would be necessary for confident direct calibration, this level of precision is valuable for the initial validation of uncalibrated models, an important tool for preliminary flood-risk analysis in settings with no gauges or rapidly changing infrastructure performance."*

I think one of the potentially important applications of the proposed method is mentioned in the discussion (lines 588-589): data for calibration of hydrodynamic models is limited, particularly for pluvial flooding. Here, estimates of water surface elevations and flood extents from cameras could fill an important data gap. If the authors could demonstrate that they can calibrate their HEC-RAS model using camera-based observations, that would strengthen the paper considerably.

While we agree with the reviewer that there is significant future opportunity in using camera-based observations to calibrate flood models (which we discuss in Lines 810-815), we feel that this effort is beyond the scope of the study presented here. The focus of the current study is application of the computer vision methods to estimate spatial flood extent, and introducing an additional model calibration element is likely to detract from that focus. The reviewer's suggestion would require us to simultaneously evaluate the performance of our camera-based methods while also applying those methods to calibrate the HEC-RAS model, introducing ambiguity into the interpretation of both elements of the study.

We will revise the discussion to more directly call out the potential use of these methods for future flood model calibration as follows:

> ***Lines 676-691****: "Camera-based observations provide a promising avenue to address these calibration gaps. Depending on the data available and the precision required, camera-derived information could support multiple levels of model calibration. At a minimum, observations of flood presence, extent, and connectivity can serve as semi-quantitative validation of model structure and behavior. More detailed or well-distributed camera installations could function as stream-gauge surrogates, enabling direct calibration of key model parameters such as surface roughness, stormwater capacity, or flood wave timing. These approaches could ultimately facilitate both post-event model evaluation and real-time model adjustment, bridging gaps in empirical data for urban flood forecasting.*
>
> *When possible to implement, camera-derived WSEs offer a rare empirical reference for validating modelled spatiotemporal patterns of inundation. For example, these high-resolution, time-resolved observations allow for direct comparison with outputs from an uncalibrated HEC-RAS Rain-on-Grid simulation of the July 4 flood event, revealing a close match in peak flood depth, timing, and extent. This proof-of-concept highlights the strong potential of integrating image-derived data into calibration workflows for 2D hydrodynamic models, particularly in high-flow scenarios where floodwaters are hydraulically connected and drainage networks are overwhelmed. Beyond event reconstruction, such observations can support real-time model updating, performance evaluation of stormwater infrastructure, and planning for flood mitigation in poorly instrumented or rapidly evolving urban settings, providing a practical, data-driven way to reduce uncertainty in urban flood simulations."*

2. Extension of flood extends beyond the camera field of view

The authors apply a flood fill procedure to estimate flood extents outside the camera field of view. I question this approach, which can't account for overland flow dynamics, infiltration, etc. I think the authors might be better off using the flood fill approach only within the field of view.

[Q5] We thank the reviewer for their comment. However, there is an important conceptual distinction between flood-fill models and process-based flood models. By prescribing a surface water level the flood-fill extents are agnostic to the factors such as infiltration that produced the water level. Further, flood fill models do not contain an explicit time-dependence, each flood extent is independently generated from a single image observation. Within the context of pluvial

flooding within an urban area, where the fraction of impervious surfaces is quite high and the study area contains many small, internally draining depression, a flood-fill model is a suitable approximation for extrapolation over short (~100m) distances.

Specific to this contribution, extrapolation serves an important role in cross-site comparison. In a low-relief environment we can expect discrete in flood connectivity, and categorical disagreements between sites reveal missing dynamics, namely storm-water infrastructure. The main point is that using flood fill models to propagate flood extents from multiple cameras improves our ability to identify these dynamics.

Given the widespread use of static water-level models for rapid flood assessment, it is valuable to discuss their behavior and limitations in the context of emerging flood data sources such as cameras (Gallien et al., 2014; Hong et al., 2024; Li et al., 2022; Preisser et al., 2022). We have expanded the discussion to highlight the limitations of the flood-fill approach and suggest avenues for integrating data from multiple cameras in future work.

> **Lines 612-618**: *"Flood fill methods are well-suited for short duration pluvial events in low relief, urban areas. Because the study sites within a self-contained depression, it is unlikely that there are substantial gradients in water surface elevation. This is supported by the 2D model results which, exclusive of edge effects, predicts a difference in water elevation between Sites A and B of only 0.5 cm after initial merging of the flood patches. Static elevation-based methods are widely used for rapid flood mapping, including in emergency management contexts (Gallien et al., 2024; Hong et al., 2024; Wang et al., 2024; Williams et al., 2019).. The cross-camera comparison used in this study is an effective tool for identifying potential failure modes within these models."*

3. Validation of the new method

The study would also be strengthened if estimated water surface elevations could be validated using other data sources. I understand that depth measurements may not be available, but could the authors estimate depth at strategic locations based on visible markers and compare those to estimates from their approach for the corresponding location? Also, some expanded discussion of uncertainty as a function of distance from the camera location would be beneficial.

Independent depth measurements are not available for the study site, and indeed the lack of such data is the primary motivation behind this project. However, there are identifiable markers of discrete jumps in water level, such as road overflow points (Vandele et al. 2019). Based on the lidar DTMs, and aerial imagery we compared projected flood extents, with road elevation profiles above and below spillover. This is included in a revised supplementary Figure:

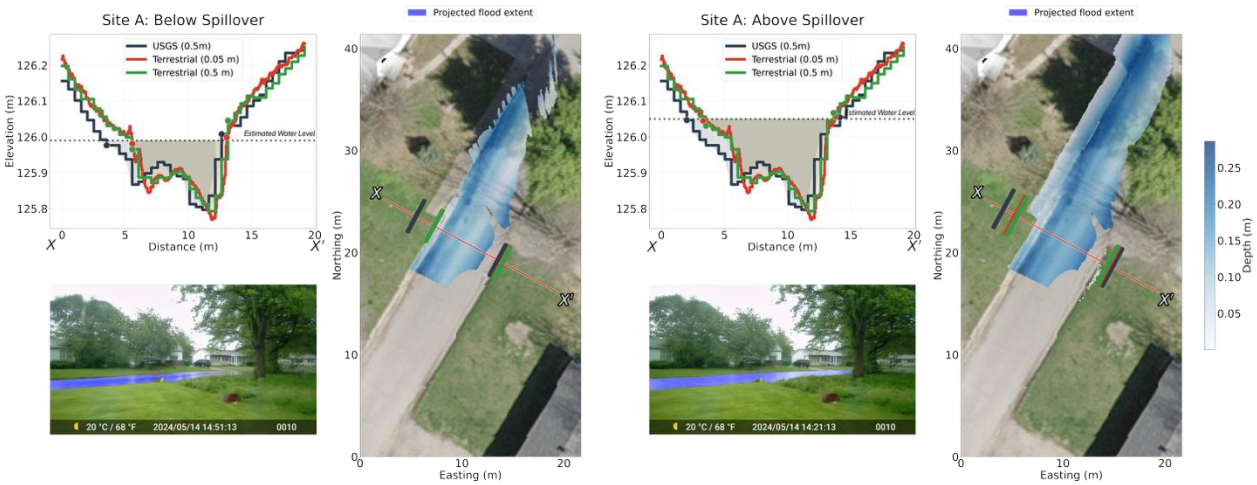

Figure S5: Road topographic profiles based on the 0.5m USGS DTM (black), 0.005m terrestrial DTM (red), and 0.5m terrestrial DTM (green). Flood mask overlays are shown with the corresponding flood extent interpolated from the mask intersection with the projected terrestrial lidar points.

"To qualitatively validate the water level extraction method we examined observation immediately above and below overtopping of the road boundary. These were compared against cross-road elevation profiles extracted from both USGS and terrestrial DTMs. Prior to spillover, the projected flood extent ends at the road boundary, with water level slightly below the curb elevation found in the terrestrial DTM profile. After spillover, the project extent expands to fill the small paved area above the road surface, before stopping at the lawn boundary. This is consistent with the image observation, and topographic profile of a second spillover onto the lawn itself. Further from camera, toward the NW, decreased pixel resolution leads to the projected extent bleeding beyond the road, potentially upwardly biasing estimated water levels. The extracted water level is approximately 3 cm higher than the elevation contour best aligned with the projected flood boundary below spillover, and approximately 1cm higher after spillover. The magnitude of both these biases decreases with larger flood extents due to more gradual elevation gradients, and the lack of curb shadows."

We have added discussion of distance dependent pixel resolution and its influence on camera pose, and water level estimation:

Line 318-331: "*Because pixel resolution decreases with distance from the camera, multiple 3D points may project onto a single image pixel; therefore, all inundated points are retained and a one-to-one pixel–point correspondence is not enforced. Together, these inundated points represent the portion of the ground surface that is underwater at the time the image was captured. This set of inundated points is interpolated into a 0.05-meter resolution raster representing the visible flood extent in the image. This interpolation step reduces bias associated with distance-dependent differences in point density and avoids over-representation of regions where many 3D points project to a single pixel. Water surface elevation (WSE) is estimated from the rasterized flood boundary rather than from individual pixels or raw point projections. Canny edge detection is applied to the rasterized inundation extent to identify the flood boundary, and the 90th and 95th percentiles of the resulting edge elevation distribution (WSE_{90} and WSE_{95}) to account for potential topographic noise or obstruction of the water edge in the time lapse images. Assuming a flat water surface, elevations along the flood boundary should exhibit a sharp peak at the upper end of the elevation distribution. The consistency and sharpness of this peak are another parameter useful to evaluate the camera pose estimation, as errors in estimated camera orientation or translation produce unrealistically large elevation differences between near- and far-field water edges."*

Other comments

Figure 1b – this image is difficult to interpret, perhaps change the color scheme?

[Q7] We adjusted this figure to accentuate the small-scale variation of the floodplain, while preserving the context of the upland bluffs.

Section 2.3 – I recommend revising this section. It is difficult to follow the detailed accounts of start and end times. It might be better to display this as a figure or omit some of the detail not necessary for understanding the larger picture.

[Q8] We will add additional annotation of key elements of flood timing and duration to figures 3 & 4 and will remove unnecessary details from the text. However, we feel that some narrative description of the events as observed by the cameras is necessary to build reader intuition and understanding and better prime them for presentation of the flood extent estimates.

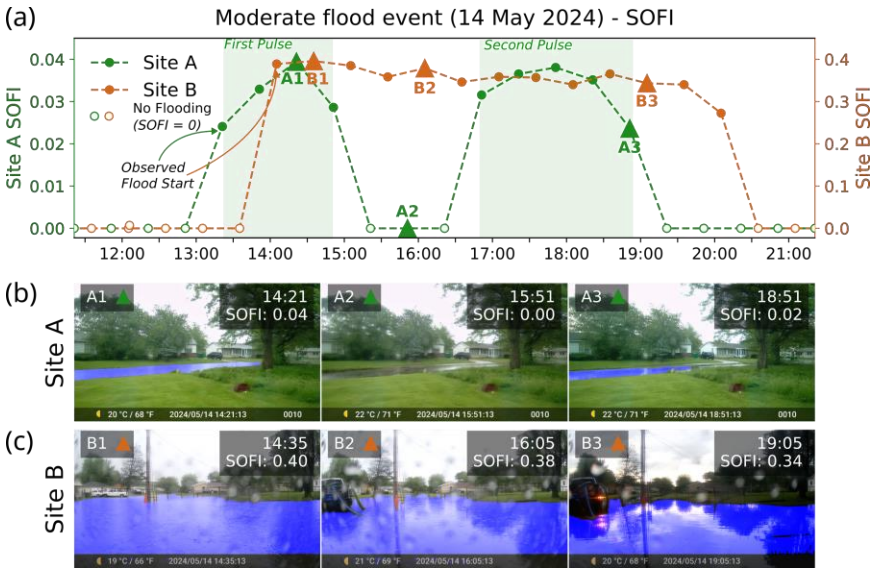

Figure 3: (a) SOFI time series for 14 May moderate severity case study event. Representative flooded images from (b) Site A and (c) Site B. Segmented flood masks are shown in blue.

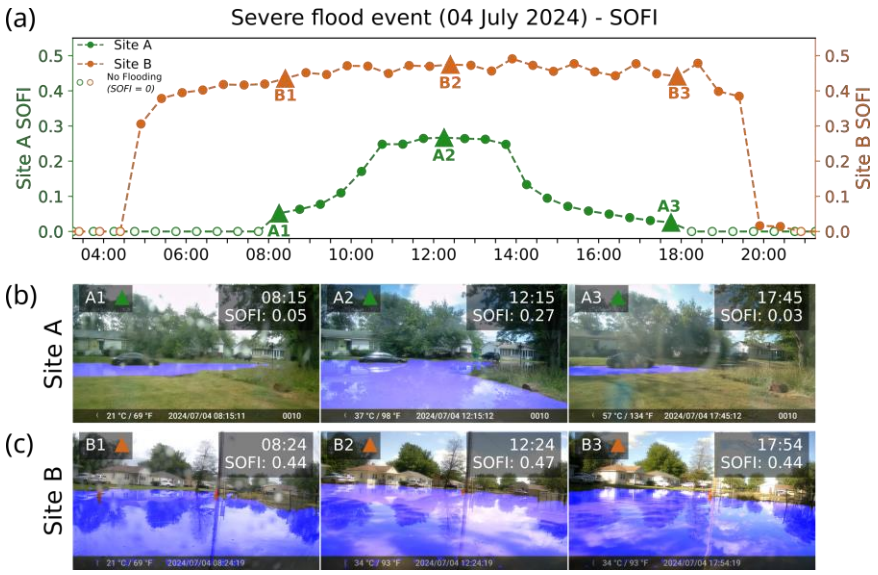

Figure 4: (a) SOFI time series for 04 July severe case study event. Representative flooded images from (b) Site A and (c) Site B. Segmented flood masks are shown in blue.

Line 322: What do the authors mean by “rainfall was uniformly applied to the domain"? Precipitation data from one gauge was applied to the entire model domain? Uniform intensity? Please clarify.

[Q9] A single rain gauge record was used for each flood event, and that hyetograph was applied to the entire model domain grid, with no spatial variability in precipitation. While a simplification, the dominance of local pluvial runoff in the study area, and the short duration of the case-study flood events, likely limits the influence of watershed scale precipitation gradients, and would not alter the basic behaviors of the rain-on-grid model relevant for comparison to the camera-derived estimates. We have clarified this in the methods:

> ***Lines 347-349:*** *“This model is implemented using the Hydrologic Engineering Center’s River Analysis System (HEC-RAS), configured with a “rain-on-grid” unsteady boundary condition to simulate overland water flow across an 89.6 km² model domain covering the study site (USACE, 2022). The base terrain is the 0.5 m USGS DTM. Rainfall records defined the unsteady inputs the model domain, assuming spatially uniform precipitation.”*

Line 365: Please explain how SOFI values should be interpreted.

[Q10] SOFI is the fraction of the total fraction of an image classified as flooded. It is included as a semi-quantitative metric of flood magnitude to contextualize the estimated water level, and flood extent trends. However, the absolute values of SOFI are a function of both the physical flood extent, and the perspective of the camera. For example, a camera installed directly Infront of a flood source will see SOFI initially rise very quickly, before leveling off as flooding fills the FOV (see SI Figure 2). We have added additional text explaining this interpretation:

Line 250-254: *"This ratio is referred to as the Static Observer Flooding Index (SOFI), following the approach of Vitry et al. (2019), providing a simple proxy for flood intensity as seen from a fixed observation point. SOFI has been shown to correlate strongly with changes in water level for a given location (Moy de Vitry et al. 2019). The shape and magnitude of SOFI response depend strongly on the geometry of a camera relative flooding, and as such values cannot be directly compared between study sites."*

'Comment on egusphere-2025-3962', Anonymous Referee #3, 28 Dec 2025

The paper is very interesting and presents an application of camera-based flood monitoring. I do not fully agree with some of the terminology used to describe existing approaches, particularly the use of the term "traditional," but this does not significantly affect the main message of the paper. Given the rapid improvement in machine learning methods, camera quality, and the continuously decreasing costs of imaging systems, I see strong potential for camera-based approaches in the near future. The proposed methodology is timely and well aligned with these developments. Overall, I think the paper can be considered for publication after the authors address the comments listed below and other reviewers' comments.

We thank the reviewer for their thoughtful comments on our manuscript. We answer any questions and describe any changes made to the text below each comment below.

1- The statement that traditional fluvial monitoring infrastructure (e.g., stream gauges and water level sensors) is not suited to detect spatially disconnected pluvial flood patches appears overstated. The manuscript should acknowledge that recent advances in dense sensor networks, smart drainage monitoring, and urban hydrometric instrumentation partially address these limitations. The authors are encouraged to moderate the claim and more clearly delineate the *specific contexts* (e.g., highly localized, shallow, short-duration inundation) in which conventional monitoring remains insufficient.

2- Lines ~(75–90): The manuscript currently frames alternative approaches in a way that may be interpreted as dismissive of prior work. It is not necessary to portray existing methods as fundamentally inadequate to justify the proposed approach. A more balanced

framing that highlights complementary strengths and weaknesses particularly regarding accuracy, spatial resolution, temporal resolution, and operational constraints would strengthen the motivation and credibility of the study.

We thank the reviewer for these comments. We have added additional context on other state of the art monitoring techniques to better highlight the specific applications where camera-based monitoring has advantages, both in terms of flooding process, and implementation/operational considerations (Mydlarz et al. 2024; Silverman et al. 2024; Gold et al 2023). We have also modified the language to avoid appearing dismissive of other approaches but still highlight where camera-based approaches can build new capacity.

> ***Lines 81-86: "****Both contact sensors (e.g., pressure transducers) and non-contact sensors (e.g., radar, ultrasonic), have proven effective for monitoring distributed urban flooding (Mydlarz et al. 2024; Gold et al. 2023). However, they can face operational challenges for small-scale flooding in urban settings, including limited installation locations and sensitivity to local disturbances (Song et al., 2024). For example, the radar based FloodNet system was limited at many sites to placement over sidewalks, limiting observation of early road flooding (Mydlarz et al. 2024). "*
>
> ***Lines: 718-732: "****Cameras offer a flexible, low-cost, and highly informative option for distributed monitoring in settings where flooding dynamics are poorly understood. However, even under ideal conditions image-based water level estimates are unlikely to reach the absolute accuracy of pressure-based sensors, and the sensitivity of image quality to environmental conditions make them less well suited for contexts where the consistency of measurement in paramount. Particularly when more direct public sector cooperation is possible, storm drain installed pressure sensors have proven highly valuable for realtime distributed flood monitoring (Gold et al. 2024; Silverman et al. 2024). However, even in these cases, cameras can serve as important complement to non-visual sensors. Gold et al. 2024 used co-located cameras and storm-drain pressure sensors, with images helping to identify cases where storm drain based measurements may be impaired. Another major advantage of semi-permanent cameras compared to other sources of flood images, such as public webcams, security cameras, or crowdsourced photos, is the flexibility to adapt the network while otherwise maintaining stability in observations (Helmrich et al., 2021). While camera sensors themselves are highly available, the requirement of high-resolution topographic data can still be a barrier to broader application of our methods. However, advancements such as smartphone mounted lidar, and national scale datasets like 3DEP have helped to mitigate these challenges. Future research could leverage this framework to optimize camera network configuration, balancing the number and placement of ground-based cameras to maximize spatial coverage and the ability to observe flood connectivity (Negri et al., 2025; Zhao et al., 2025)."*

3- The challenge of translating two-dimensional image-based flood fractions into real-world water depth should be explained more rigorously. In heterogeneous urban environments, flooded pixel fraction does not scale linearly with water depth because inundation often occurs in shallow, spatially discontinuous depressions controlled by microtopography, curbs, and drainage infrastructure. Small vertical changes in water level can produce large apparent changes in flooded area (or vice versa), leading to ambiguity when inferring depth or volume from image coverage alone. Explicitly linking this limitation

to urban surface complexity would clarify why image-only approaches are insufficient for depth estimation.

In the revised discussion we draw more explicit contrast between the image-only SOFI characterization of flooding, and the 3D-2D projection-based approach. Inundated area shows a nonlinear increase in the presence of depressions (Figure 7), with further bias in image coverage produced by camera scene geometry (SI Figure 3). While still subject to inherent resolution and visibility limitations, 3D-2D projection can overcome this limitation by explicitly accounting for the relative geometry of flood extents to the camera:

> ***Lines 250-254:*** *"This ratio is referred to as the Static Observer Flooding Index (SOFI), following the approach of Vitry et al. (2019), providing a simple proxy for flood intensity as seen from a fixed observation point. SOFI has been shown to correlate strongly with changes in water level for a given location (Moy de Vitry et al. 2019). The shape and magnitude of SOFI response depend strongly on the geometry of a camera relative flooding, and as such values cannot be directly compared between study sites."*
>
> ***Lines 550-551:*** *"Accordingly, SOFI should be interpreted primarily as a scene-specific indicator of relative change in water levels over time, rather than a measure of absolute flood magnitude or spatial extent."*

4-Use of the term "traditional" The repeated use of the term *traditional* to describe existing monitoring and modeling approaches is potentially misleading. Many of these methods are actively evolving and increasingly integrated with high-resolution data and advanced numerical schemes. The authors may consider replacing this term with more precise language to avoid implying obsolescence.

In revision we removed use of the term 'traditional' throughout and instead refer to either specific technologies or specific distinguishing features, such as between contact/point-based and non-contact/continuous methods.

5-Is this method having a lower cost in compare with existing approaches? this argument is weakened by reliance on aerial LiDAR and high-density terrestrial LiDAR. Why the authors should not explicitly discuss the cost, accessibility, and transferability of these datasets, particularly for low-income or data-scarce regions. A comparative table summarising data requirements, costs, spatial/temporal resolution, and uncertainties across camera-based methods, LiDAR-dependent approaches, and conventional monitoring would provide a transparent and unbiased comparison.

We have moderated the language to instead discuss cameras as one member of a suite of distributed, low-cost monitoring tools. However, we emphasize that the aim of this

study is not a ready-made analysis package or hardware platform and feel that a specific cost-breakdown comparison of our deployment may not be representative of camera monitoring as a whole. And given the significant project and site-specific factors involved in sensor network cost feel that explicit cost comparison would be a diversion from the main goal of demonstrating our methodology. We have however added discussion of tradeoffs between resolution, accuracy and cost/flexibility between methods. We recognize that there are still material and data barriers to applying some of our methods. However, we note that most elements of our workflow could be replicated in less intensive ways, albeit with scale and accuracy tradeoffs. For example, at the scale of an individual camera recent studies have leveraged smartphone integrated lidar and photogrammetry for water level prediction (Erfani et al. 2023), while the aerial lidar was freely obtained from the USGS 3DEP program. (Refer to Lines 854-858 in our reply to Questions 1 & 2)

6-The workflow for estimating floodwater elevation is central to the contribution of this paper, yet it is difficult to fully evaluate due to incomplete access to it (the Zenodo link could not be accessed, at least I could not).

The current repository is still in 'draft' mode to facilitate any material additions or changes prompted by the review process. The sharing link in the Code Availability section (reproduced below) should give access and we are happy to generate and share a new link if needed.

https://zenodo.org/records/16414887?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjAzM2MxZjBhLTJjNWEtNDI1MS04ZWE1LTRlODJlZTkzMjEyNCIsImRhdGEiOnt9LCJyYW5kb20iOiJhNmVlNDQ0Y2Y0Njc3MTFiZDQ0MzAzMGI5ZDFmYmNkOSJ9.fNd50BKqMzWA7NBgVwrWqpGVKyLTJFSjcn_yawfLlnX3YDGyoL1NwX4-qnnuGKgT6coHGLrntXJGKay-RhatKw

7-The Discussion and Conclusions section is lengthy and combines interpretation with summary statements. I recommend separating this into a concise Discussion section focused on interpretation and limitations, followed by a distinct Conclusions section that succinctly highlights the main contributions, findings, and implications.

We will add subheadings to the discussion and separate the conclusions into a separate and final section.

8-I understand the study is not centred on HEC-RAS, modelling, however, providing a brief description of the HEC-RAS setup, assumptions, and any calibration or validation strategy

(ideally in supplementary material) would strengthen the credibility of the comparison without disrupting the narrative flow of the main manuscript.

We have migrated details regarding HEC-RAS model design and execution from the code supplement into the main text and supplement (See Supplementary Text 1). We have also added additional explanation of the motivation, and limitations of the current model comparison:

> ***Lines 370-373:*** *"Because the model itself is only qualitatively calibrated, its output is not treated as a direct validation for absolute water levels estimated from images. Instead, it characterizes similarity or divergence in flood behavior predicted by each method, based. This is quantified both in terms of the relative agreement in predicted flood extent, and spatial flood connectivity, between the two methods."*

> ***Lines 701-709: "****Despite these challenges, our results demonstrate how empirically-derived WSEs can complement and strengthen traditional hydraulic modelling workflows. Our method provides continuous, high-resolution estimates of water level and extent that are directly tied to real flood behaviour, capturing sub-decimeter changes in WSE and floodwater connectivity that would otherwise be missed by point-based flood monitoring and modelling approaches. While further validation of camera-derived extents would be necessary for direct model calibration, this level of precision is valuable for the initial validation of uncalibrated models, an important tool for preliminary flood-risk analysis in settings with no gauges or rapidly changing infrastructure performance. As stormwater systems become increasingly strained by climate extremes, integrating data-driven camera networks with physically-based modeling frameworks offers a promising pathway for improving urban flood forecasting, response, and planning. "*

9-The manuscript would benefit from a brief discussion of how segmentation uncertainty propagates into water surface elevation estimates, particularly under challenging conditions such as specular reflections, shadows, low-light conditions, and partial occlusions by vegetation or vehicles. (At least raise them)

This point is well taken and an important aspect of the broader viability of these methods. This is partly included in the discussion of trade-offs of camera-based monitoring in our response to question two. We will also add discussion of both random and systematic segmentation error, noting the ways that our interpolation approach partly reduces their influence:

> ***Lines 641-660:***
>
> *Elements of our method – including spatial aggregation of flood boundaries and the calculation of multiple water-level thresholds – effectively constrain uncertainty to levels suitable for urban flood characterization. We find that water-level estimates are robust to both minor random and systematic errors in flood-mask segmentation. For the severe event at Site A, water levels calculated for the half of the scene containing a parked car that partially obscured the water line differed by less than 2 cm, on average, from estimates derived from the unobstructed portion of the scene. Similarly, introducing random jitter of 10–20 pixels to*

the flood-mask boundary produced mean water-level differences of less than 2 cm. More substantial errors in flood segmentation or camera pose that are not mitigated by the method are typically identifiable through diagnostic artifacts, including large reprojection errors, asymmetric projected flood extents, or exaggerated differences between WSE90 and WSE95. In addition, the visual context provided by the images allows qualitative validation against observable flood indicators such as roadway overtopping.

Beyond water-level uncertainty, flood-extent estimates are influenced by the quality of the underlying topographic data. Even with careful georeferencing, physical landscape change between surveys or differences in lidar point density can introduce localized elevation discrepancies. Flood extents propagated using aerial lidar tend to be biased toward overprediction because fine-scale topographic structures, such as curbs or drainage ditches, are only partially resolved. Both water-level estimation and flood-extent propagation may therefore be most sensitive at lower water levels, where small-scale topographic features exert stronger control on flood extent. Consequently, interpretations of discrete changes in flood connectivity resulting from small increases in water level should be treated cautiously. However, because the camera images directly capture the spatial distribution of floodwaters between sites, they can provide an independent observational check on modelled flood connectivity and allow clear identification of locations where modelled inundation diverges from observed flooding. Future work should further investigate how uncertainty propagates among these sources.

Reply Citations:

Gallien, T. W., Sanders, B. F., & Flick, R. E. (2014). Urban coastal flood prediction: Integrating wave overtopping, flood defenses and drainage. Coastal Engineering, 91, 18-28.

Gold, A., Anarde, K., Grimley, L., Neve, R., Srebnik, E. R., Thelen, T., ... & Hino, M. (2023). Data from the drain: A sensor framework that captures multiple drivers of chronic coastal floods. Water Resources Research, 59(4), e2022WR032392.

Hong, Y., Kessler, J., Titze, D., Yang, Q., Shen, X., & Anderson, E. J. (2024). Towards efficient coastal flood modeling: A comparative assessment of bathtub, extended hydrodynamic, and total water level approaches. Ocean Dynamics, 74(5), 391-405.

Li, Z., Mount, J., & Demir, I. (2022). Accounting for uncertainty in real-time flood inundation mapping using HAND model: Iowa case study. Natural Hazards, 112(1), 977-1004.

Negri, R., Ceferino, L., & Cremen, G. (2025). Prioritizing urban areas for the deployment of hyperlocal flood sensors using stakeholder elicitation and risk analysis. Natural Hazards Review, 26(3), 04025020.

Silverman, A. I., Brain, T., Branco, B., sai venkat Challagonda, P., Choi, P., Fischman, R., ... & Toledo-Crow, R. (2022). Making waves: Uses of real-time, hyperlocal flood sensor data for emergency management, resiliency planning, and flood impact mitigation. *Water Research*, *220*, 118648.

Williams, L. L., & Lück-Vogel, M.: Comparative assessment of the GIS based bathtub model and an enhanced bathtub model for coastal inundation. Journal of Coastal Conservation, 24(2), 23, 2020.

Zhao, Z., Liang, Y., Wang, K., Ding, X., Zhang, Y., and Hu, C.: Collaborative sensing optimization layout model of heterogeneous sensors under urban flooding environment. Journal of Hydrology, 650, 132528, 2025.