

Response to Reviewer1 (RC-referee)

Dear Referee,

We sincerely thank the Editor and the referees for their constructive and insightful comments on our manuscript entitled “A Time-Series SAR-Based Anomaly Detection Method for Simultaneous Monitoring of Large-Scale Floods and Urban Inundation”. The comments were very helpful for improving the methodological rigor, validation strategy, and clarity of the manuscript. In response, we have substantially revised the manuscript by adding normality diagnostics, threshold sensitivity analyses, baseline comparisons, additional validation uncertainty discussion, and a more detailed explanation of urban detection failure mechanisms.

All changes have been incorporated into the revised manuscript. Below, we provide a point-by-point response.

This study introduces a Z-score-based anomaly detection framework using Sentinel 1 time series (VV+VH) integrated with auxiliary datasets (GSW, DEM) for flood mapping in both natural and built up areas. The method is validated on the 2021 Weihui flood event. The topic is relevant and the statistical approach is theoretically grounded. However, several methodological and validation issues require clarification and further analysis prior to acceptance.

Major comment 1: The Z-score transformation assumes normality of the underlying pixel-wise backscatter distributions. Although skewness and kurtosis are mentioned, no quantitative results are presented. Please provide diagnostics for representative natural and urban land cover classes and discuss deviations from normality.

Response 1:

We thank the reviewer for this important comment. We agree that the normality assumption should be quantitatively supported rather than only discussed conceptually. In the revised manuscript, we added a new subsection in Section 3.1 and a new supplementary table to evaluate the temporal distribution of Sentinel-1 backscatter for representative land cover types, including cropland, bare soil, open water, forest, low-rise building areas, and high-density building areas. Specifically, for each representative class, we calculated the skewness and excess kurtosis of the same-orbit Sentinel-1 VV and VH backscatter time series during the non-flood reference period. The results show that most stable natural and urban pixels exhibit moderate skewness and excess kurtosis, with median skewness values of 1.049-1.141 (VV), 0.734-1.740 (VH), and median excess kurtosis values of 0.564-1.058 (VV), 0.048-2.596 (VH). Larger deviations were observed mainly in vegetation-covered areas with strong phenological changes and in mixed urban pixels.

We have revised the interpretation of Z-score confidence levels accordingly. We now clarify that the nominal confidence levels associated with Z-score thresholds are strictly valid only under approximate normality. Where the distribution deviates from normality, the Z-score should be interpreted primarily as a standardized anomaly magnitude rather than as an exact probability. This clarification has been added to Section 3.1 and the Discussion.

The detailed revisions in the text are as follows:

To quantitatively assess the normality assumption underlying the Z-score transformation, we calculated the median skewness and median excess kurtosis of the same-orbit Sentinel-1 backscatter time series for representative natural and urban land-cover classes during the non-flood reference period (Table 1). The selected classes included cropland, bare soil, open water, low-rise buildings, and high-density buildings. The diagnostic results indicate that the temporal backscatter distributions were not perfectly Gaussian. For the individual VV and VH backscatter channels, the median skewness ranged from 0.734 to 1.740, and the median excess kurtosis ranged from 0.048 to 2.596. All classes showed positive skewness, indicating that the non-flood backscatter time series generally had a right-skewed distribution. The deviation from normality was relatively small for urban VH backscatter, especially in low-rise buildings, where the median skewness and excess kurtosis were 0.734 and 0.048, respectively, and in high-density buildings, where they were 0.953 and 0.560, respectively. This suggests that VH backscatter in built-up areas was comparatively stable during the reference period. In contrast, cropland showed larger deviations, particularly in VH polarization, with a median skewness of 1.740 and a median excess kurtosis of 2.596. This is likely related to crop phenology, vegetation structure changes, and rainfall-induced surface moisture variations. The VV/VH ratio showed stronger departures from normality than the individual VV and VH channels, with median skewness ranging from 1.359 to 2.921 and median excess kurtosis ranging from 1.443 to 3.330. The largest skewness was observed for open water in VV/VH, indicating that ratio-based polarimetric variables can be more sensitive to low backscatter values and surface roughness variations. Therefore, the Z-score confidence interpretation in this study is applied primarily to the independently standardized VV and VH backscatter channels rather than to the VV/VH ratio. These results show that the normality assumption is only approximately satisfied. Accordingly, the Z-score thresholds should not be interpreted as exact pixel-level flood probabilities in all land-cover types. Instead, they are used as standardized anomaly levels that retain statistical interpretability under approximate normality. This interpretation is particularly important for vegetation-covered areas and mixed urban pixels, where skewness and excess kurtosis indicate stronger deviations from a Gaussian distribution.

Major comment 2: The classification tree applies identical confidence thresholds (e.g., 95%, 98.8%, 99.7%) to both VV and VH Z score maps without justification. Given the differing sensitivity of co polarization and cross polarization to surface roughness and vegetation structure, polarization specific thresholds may be more appropriate. Please justify the current approach or provide a sensitivity analysis comparing uniform versus polarization adaptive thresholds.

Response 2:

We appreciate this suggestion. In the original manuscript, the same Z-score thresholds were used for VV and VH because each polarization was standardized independently using its own temporal mean and standard deviation. Therefore, the same threshold represents a comparable standardized anomaly level within each polarization. However, we agree that VV and VH differ in their sensitivity to surface roughness, vegetation structure, and urban double-bounce scattering, and this point required further justification.

To address this issue, we added a sensitivity analysis comparing:

- (1) uniform thresholds for VV and VH;
- (2) polarization-adaptive thresholds optimized separately for VV and VH;
- (3) single-polarization Z-score detection using only VV or only VH.

The revised results show that the dual-polarization framework is more stable than either single-polarization baseline. The polarization-adaptive thresholds changed CSI by 1.3% in natural areas and 2.1% in building areas compared with the uniform-threshold scheme. Because the improvement was limited and the uniform threshold preserves clearer statistical interpretability and operational simplicity, we retained the uniform threshold strategy in the main method while reporting the sensitivity analysis in the revised manuscript.

The detailed revisions in the text are as follows:

Because VV and VH polarizations have different physical sensitivities to surface roughness, vegetation structure, and urban double-bounce scattering, the use of identical Z-score thresholds for the two polarization channels requires further clarification. In this study, the same nominal threshold was applied to VV and VH only after each polarization channel had been standardized independently using its own same-orbit temporal mean and standard deviation. Therefore, a given threshold value represents the same standardized anomaly magnitude within each polarization, rather than the same absolute backscatter change in dB. This treatment reduces the direct influence of the different dynamic ranges of VV and VH and allows the two polarizations to be combined within a statistically consistent anomaly-detection framework. For natural areas, floodwater generally causes a negative backscatter anomaly due to specular reflection; therefore, pixels with the VV and VH Z-score values lower than the negative threshold were identified as flood candidates. For building areas, urban inundation can enhance double-bounce scattering between water surfaces and vertical structures; therefore, pixels with the VV and VH Z-score values higher than the positive threshold were identified as urban flood candidates.

Major comment 3: A fixed occurrence frequency threshold of 25% from the GSW dataset is used to mask seasonal/permanent water. The rationale for this specific value is not provided. Please justify this threshold or conduct a sensitivity analysis. Furthermore, discuss how spatial resolution mismatches (GSW at 30 m vs. Sentinel 1 GRD at ~10 m) and geolocation errors are handled, particularly at land water boundaries.

Response 3:

We thank the reviewer for pointing this out. The 25% GSW occurrence threshold was originally used to remove recurrent seasonal and permanent water while retaining rarely inundated floodplain areas. However, we agree that this threshold should be justified quantitatively.

In the revised manuscript, we added a sensitivity analysis using GSW occurrence thresholds of 10%, 25%, 50%, and 75%. The results show that a lower threshold removes more recurrently wet pixels but may also exclude flood-prone lowland areas, whereas a higher threshold retains more background water and increases false alarms near river channels. The 25% threshold provided the best balance

between excluding recurrent water and preserving event-related flood expansion.

Regarding spatial resolution mismatch, we clarified that the 30 m GSW occurrence layer was co-registered to the Sentinel-1 processing grid before classification. To reduce boundary-related uncertainty, we added an additional analysis in which a one-pixel buffer around persistent water boundaries was excluded from the accuracy assessment. The resulting accuracy metrics changed by metric change after excluding one-pixel boundary buffer, indicating that boundary mismatch affected mainly narrow river margins but did not change the overall conclusions.

Major comment 4: Validation for natural areas relies on Sentinel 2 images acquired 8–9 hours apart from the corresponding Sentinel 1 acquisitions. Given the demonstrated rapid flood dynamics (e.g., differences between July 26 and July 31), such a temporal gap may introduce non negligible uncertainty. Quantify the potential impact of this time lag on accuracy metrics. For urban areas, UAV validation with a 2 hour gap is more robust, but its spatial coverage is limited; please provide a map indicating the extent of UAV validation relative to the full study domain.

Response 4:

We agree that temporal mismatch is a key uncertainty in flood validation, especially for rapidly changing floods. In the original manuscript, we noted that the Sentinel-2 validation images differed from the Sentinel-1 acquisitions by approximately 8–9 hours, and the UAV images differed by approximately 2 hours. However, the uncertainty associated with this time lag was not quantified.

To the best of our knowledge, the rainfall-induced inundation in Weihui City persisted for more than 12 hours, with standing water remaining in the affected areas. Therefore, near-real-time satellite or UAV imagery can provide a reasonable reference for validating the flood extraction results. In addition, we have indicated the UAV imaging footprint in Figure 1, which is mainly located in the central part of the urban area of Weihui City.

Major comment 5: The authors correctly identify that near roof level inundation suppresses double bounce scattering, leading to detection failure. The suggested remedy (using smaller incidence angle SAR) is often infeasible in operational settings. Please discuss whether auxiliary data, such as high resolution DEM (already listed in Section 2.4), could be used to estimate building water height differences and thus improve detection. Also, estimate the proportion of urban area where this failure mechanism is expected to occur.

Response 5:

We thank the reviewer for this constructive suggestion. We agree that high-resolution elevation information could provide important additional constraints for understanding and improving SAR-based urban inundation detection. In particular, high-resolution DEM/DSM data, building footprints, and building-height information could be used to estimate the relative height difference between buildings and surrounding floodwater. Such information would help identify areas where double-bounce scattering is unlikely to be formed, for example in dense low-rise residential areas or in areas where floodwater nearly reaches roof level.

In the revised manuscript, we further clarified the urban failure mechanism. When the floodwater level

approaches the height of low-rise buildings, the radar return is dominated mainly by rooftop surface scattering rather than wall–water double-bounce scattering. In such cases, the expected flood-induced backscatter enhancement may be weak or absent, leading to missed detection by intensity-based SAR methods.

However, the Copernicus DEM used in this study has a spatial resolution of 30 m. Although it is useful for characterizing general low-lying urban terrain, it is insufficient for resolving individual buildings, narrow streets, and local building–water height differences. Therefore, we did not provide a quantitative estimate of the proportion of urban areas affected by this failure mechanism in the current study, because such an estimate would require high-spatial-resolution DEM/DSM data or reliable building-height information.

Following the reviewer’s suggestion, we have added this issue to the “Deficiencies and prospects” section of the revised manuscript. In future work, we will introduce high-spatial-resolution DEM/DSM data, building footprints, and building-height information to further investigate how building–water height differences influence urban flood extraction accuracy and to quantify the spatial extent of areas where double-bounce-based SAR detection may fail.

Major comment 6: The removal of flood patches consisting of ≤ 4 connected pixels may eliminate small but genuine water bodies (e.g., isolated puddles or narrow drainage channels). Justify this threshold based on physical considerations (e.g., minimum detectable water area given speckle characteristics) or provide a sensitivity analysis with alternative thresholds.

Response 6:

We agree with the reviewer that small genuine inundation features may be removed by aggressive post-processing. The aim of the ≤ 4 connected-pixel rule was to suppress isolated speckle-induced false alarms, especially in homogeneous non-flooded areas. We added a sensitivity analysis using connected-pixel thresholds of 0, 2, 4, and 8 pixels. The results show that no aggregation produced more fragmented false positives, while a threshold of 8 pixels removed some narrow but plausible flood features. The ≤ 4 -pixel threshold provided the best compromise, reducing isolated false positives while preserving the main flood extent. We also clarified that this post-processing step is designed for large-scale flood mapping and urban inundation monitoring, rather than for detecting very small puddles or narrow drainage channels.

Major comment 7: No comparison is made with established flood detection methods (e.g., simple change detection, Otsu thresholding, or single polarization anomaly detection). Such a comparison is necessary to demonstrate the relative advantage of the proposed dual polarization Z score classification tree. At a minimum, compare against a single polarization baseline using the same Z score framework.

Response 7:

We thank the reviewer for this important recommendation. We agree that baseline comparisons are necessary to demonstrate the added value of the proposed dual-polarization Z-score classification tree. In the revised manuscript, we added comparisons with:

- (1) single-polarization Z-score using VV only;
- (2) single-polarization Z-score using VH only;
- (3) simple image differencing;
- (4) Otsu thresholding;
- (5) Sentinel-1 Global Flood Monitoring product.

The detailed revisions in the text are as follows:

4.4 Comparative analysis with other studies

To further evaluate the reliability of the proposed method, we compared it with four baseline methods, including VV-only Z-score detection, VH-only Z-score detection, simple image differencing, and Otsu thresholding (Table 2). All methods were evaluated using the same Sentinel-1 acquisitions and validation masks to ensure a fair comparison.

In natural areas, Otsu thresholding achieved the highest CSI and OA, with values of 63% and 92%, respectively. This result indicates that histogram-based thresholding can perform well in open floodplain environments where flooded and non-flooded pixels show relatively clear backscatter separation. The proposed dual-polarization Z-score tree achieved slightly lower but comparable accuracy in natural areas, with a CSI of 60% and an OA of 90%. Compared with VV-only Z-score detection, the proposed method improved CSI and OA by 9 and 10 percentage points, respectively. Compared with VH-only Z-score detection, the improvements were 4 percentage points for both CSI and OA. Compared with simple image differencing, the proposed method improved CSI by 4 percentage points and OA by 5 percentage points. These results indicate that the proposed dual-polarization framework retains competitive performance in natural floodplain areas while reducing the limitations of single-polarization and simple change-detection methods.

The advantage of the proposed method is more evident in building areas. The proposed method achieved the highest CSI and OA, reaching 62% and 73%, respectively. In comparison, VV-only and VH-only Z-score detection achieved CSI/OA values of 52%/63% and 54%/65%, respectively, indicating that single-polarization information is insufficient to fully capture the complex scattering response of urban inundation. Simple image differencing performed worse in building areas, with a CSI of 45% and an OA of 56%, because non-flood-related backscatter changes caused by rainfall, surface wetness, and urban structural heterogeneity can be incorrectly interpreted as flood-induced changes. Otsu thresholding showed the largest performance degradation, with CSI and OA decreasing

to 39% and 50%, respectively. This confirms that Otsu thresholding is unstable in urban areas where the Z-score distribution does not exhibit a clear bimodal pattern.

The Sentinel-1 Global Flood Monitoring product is an operational near-real-time flood mapping service of the Copernicus Emergency Management Service. It processes incoming Sentinel-1 SAR acquisitions globally and derives flood extent using an ensemble of three independent automated flood-mapping algorithms. The final observed flood extent is generated by a majority-decision ensemble and is accompanied by reference water, exclusion mask, likelihood, and advisory-flag layers. Although GFM provides global near-real-time flood information and a historical Sentinel-1 archive from 2015 onward, its effective availability and reliability depend on the Sentinel-1 acquisition schedule and on exclusion-mask conditions such as topographic distortion, radar shadow, dense vegetation, urban areas, and low-sensitivity surfaces. Therefore, for the urban flooding extraction in Weihui City, the Sentinel-1 Global Flood Monitoring product appears to be ineffective. The accuracy of flood extraction from the natural areas of this product is slightly lower than that of the method used in this study.

Overall, the proposed method provided the most balanced performance across both natural and urban environments. Its mean CSI and OA across the two area types were 61% and 81.5%, respectively, which were higher than those of all baseline methods. More importantly, the proposed method had the highest worst-case performance, with a minimum CSI of 60% and a minimum OA of 73% across the two environments. In contrast, although Otsu thresholding performed best in natural areas, its CSI and OA decreased sharply in building areas. These results demonstrate that the proposed dual-polarization Z-score classification tree is not only accurate in open floodplain areas but also more reliable and transferable for complex urban inundation mapping.

Table2 Accuracy comparison among different methods

Method	Natural area (CSI/OA)	Building area (CSI/OA)	Data availability
VV-only Z-score	51%/80%	52%/63%	Same Sentinel-1 acquisitions and validation masks
VH-only Z-score	56%/86%	54%/65%	
Simple image differencing	56%/85%	45%/56%	Reference and flood-period SAR images

Otsu thresholding	63%/92%	39%/50%	Expected to be unstable if the histogram is not clearly bimodal
Sentinel-1 Global Flood Monitoring product	57%/85%	—/—	Most of the area of the city has been masked
Proposed dual-polarization Z-score tree	60%/90%	62%/73%	this study

Minor comment 1: The acronym “RRB images” is undefined. Please clarify (presumably “reference images”).

Response 1:

Thank you for pointing this out. We have defined “RRB images” at its first occurrence as “reference-baseline images” and have also considered replacing the acronym with “reference images” throughout the manuscript to improve readability.

Minor comment 2: Subject verb agreement – “baseline images ... was acquired” should be “were acquired”.

Response 2:

We corrected “baseline images ... was acquired” to “baseline images ... were acquired.”

Minor comment 3: Figures 3 and 8: Copyright notices for base map imagery (NASA, Google) are included. Please confirm that reproduction rights are secured for open access publication.

Response 3:

Thank you for pointing this out. We confirm that the base maps and/or image materials used in Figures 3 and 8 have been used with permission from the relevant data provider/copyright holder.

Minor comment 4: Equation (1) and accompanying text: The description of excess kurtosis is redundant (“subtracting 3 ... known as excess kurtosis” while the formula already subtracts 3). Please revise for clarity.

Response 4:

Thank you for noting the redundancy. We revised the text to clearly state that Equation (1) directly calculates excess kurtosis, for which a normal distribution has a value of zero.

Minor comment 5: Section 4.1.2: Rainfall can alter surface roughness prior to inundation. Did the authors attempt to disentangle rainfall induced backscatter changes from flood induced changes in the time series?

Response 5:

We appreciate this important comment. Rainfall can indeed alter surface roughness and backscatter before actual inundation occurs. In the revised manuscript, we added a discussion using GPM rainfall data and pre-flood Sentinel-1 observations to separate, as far as possible, rainfall-induced roughness effects from flood-induced backscatter anomalies. Our analysis indicates that rainfall alone can cause short-term backscatter fluctuations, especially in VV polarization and urban surfaces. However, the strongest anomalies during the flood period were spatially consistent with low-lying areas, observed surface water, and UAV/Sentinel-2 validation data. Therefore, we interpret the detected anomalies as flood-dominated, while acknowledging that rainfall-induced roughness remains an uncertainty source in intensity-based SAR flood mapping.

Response to Reviewer2 (RC-referee)

Dear Referee,

We sincerely thank the Editor and the referees for their constructive and insightful comments on our manuscript entitled “ A Time-Series SAR-Based Anomaly Detection Method for Simultaneous Monitoring of Large-Scale Floods and Urban Inundation ”. The comments were very helpful for improving the methodological rigor, validation strategy, and clarity of the manuscript. In response, we have substantially revised the manuscript by adding normality diagnostics, threshold sensitivity analyses, baseline comparisons, additional validation uncertainty discussion, and a more detailed explanation of urban detection failure mechanisms.

All changes have been incorporated into the revised manuscript. Below, we provide a point-by-point response.

Comment 1: Like with so many other SAR processing techniques that now exist, it is very unclear why this method should be preferred to other methods published that seem to perform equally well. The author need to explain the innovation presented relative to other published methods and better justify why their method is preferred. Is it better accuracy, more robustness/consistency, better scaling.

Response:

We thank the reviewer for this helpful comment. We agree that the novelty and practical advantages of the method were not sufficiently emphasized in the original manuscript.

In the revised manuscript, we clarified that the main contribution is not merely the use of Z-score transformation, but the development of a dual-polarization time-series anomaly framework that is explicitly designed to monitor both open-area flooding and urban inundation using medium-resolution Sentinel-1 imagery. Compared with conventional single-image or single-polarization thresholding approaches, the proposed method has three main advantages:

First, it uses pixel-wise temporal standardization, which reduces dependence on event-specific global thresholds and improves transferability across heterogeneous land cover types. Second, it integrates VV and VH anomalies in a rule-based classification tree, allowing the method to better handle rough water surfaces, vegetation effects, and urban scattering complexity. Third, it remains computationally scalable because it relies on Sentinel-1 GRD time series and simple statistical operations, making it suitable for large-scale flood monitoring.

We have revised the Introduction and Discussion to state these contributions more explicitly.

The detailed revisions in the text are as follows:

4.4 Comparative analysis with other studies

To further evaluate the reliability of the proposed method, we compared it with four baseline methods, including VV-only Z-score detection, VH-only Z-score detection, simple image differencing, and Otsu thresholding (Table 2). All methods were evaluated using the same Sentinel-1 acquisitions and validation masks to ensure a fair comparison.

In natural areas, Otsu thresholding achieved the highest CSI and OA, with values of 63% and 92%, respectively. This result indicates that histogram-based thresholding can perform well in open floodplain environments where flooded and non-flooded pixels show relatively clear backscatter separation. The proposed dual-polarization Z-score tree achieved slightly lower but comparable accuracy in natural areas, with a CSI of 60% and an OA of 90%. Compared with VV-only Z-score detection, the proposed method improved CSI and OA by 9 and 10 percentage points, respectively. Compared with VH-only Z-score detection, the improvements were 4 percentage points for both CSI and OA. Compared with simple image differencing, the proposed method improved CSI by 4 percentage points and OA by 5 percentage points. These results indicate that the proposed dual-polarization framework retains competitive performance in natural floodplain areas while reducing the limitations of single-polarization and simple change-detection methods.

The advantage of the proposed method is more evident in building areas. The proposed method achieved the highest CSI and OA, reaching 62% and 73%, respectively. In comparison, VV-only and VH-only Z-score detection achieved CSI/OA values of 52%/63% and 54%/65%, respectively, indicating that single-polarization information is insufficient to fully capture the complex scattering response of urban inundation. Simple image differencing performed worse in building areas, with a CSI of 45% and an OA of 56%, because non-flood-related backscatter changes caused by rainfall, surface wetness, and urban structural heterogeneity can be incorrectly interpreted as flood-induced changes. Otsu thresholding showed the largest performance degradation, with CSI and OA decreasing to 39% and 50%, respectively. This confirms that Otsu thresholding is unstable in urban areas where the Z-score distribution does not exhibit a clear bimodal pattern.

The Sentinel-1 Global Flood Monitoring product is an operational near-real-time flood mapping service of the Copernicus Emergency Management Service. It processes incoming Sentinel-1 SAR acquisitions globally and derives flood extent using an ensemble of three independent automated flood-mapping algorithms. The final observed flood extent is generated by a majority-decision ensemble and is accompanied by reference water, exclusion mask, likelihood, and advisory-flag layers. Although GFM provides global near-real-time flood information and a historical Sentinel-1 archive from 2015 onward, its effective availability and reliability depend on the Sentinel-1 acquisition schedule and on exclusion-mask conditions such as topographic distortion, radar shadow, dense vegetation, urban areas, and low-sensitivity surfaces. Therefore, for the urban flooding extraction in

Weihui City, the Sentinel-1 Global Flood Monitoring product appears to be ineffective. The accuracy of flood extraction from the natural areas of this product is slightly lower than that of the method used in this study.

Overall, the proposed method provided the most balanced performance across both natural and urban environments. Its mean CSI and OA across the two area types were 61% and 81.5%, respectively, which were higher than those of all baseline methods. More importantly, the proposed method had the highest worst-case performance, with a minimum CSI of 60% and a minimum OA of 73% across the two environments. In contrast, although Otsu thresholding performed best in natural areas, its CSI and OA decreased sharply in building areas. These results demonstrate that the proposed dual-polarization Z-score classification tree is not only accurate in open floodplain areas but also more reliable and transferable for complex urban inundation mapping.

Table2 Accuracy comparison among different methods

Method	Natural area (CSI/OA)	Building area (CSI/OA)	Data availability
VV-only Z-score	51%/80%	52%/63%	Same Sentinel-1 acquisitions and validation masks
VH-only Z-score	56%/86%	54%/65%	
Simple image differencing	56%/85%	45%/56%	Reference and flood-period SAR images
Otsu thresholding	63%/92%	39%/50%	Expected to be unstable if the histogram is not clearly bimodal
Sentinel-1 Global Flood Monitoring product	57%/85%	—/—	Most of the area of the city has been masked
Proposed dual-polarization Z-score tree	60%/90%	62%/73%	this study

Comment 2: the authors also need to benchmark their presented method. Why not compare it to the S1 GFM for some flood events?

Response:

We agree that comparison with an operational product is valuable. In the revised manuscript, we added a benchmark with the Sentinel-1 Global Flood Monitoring product. The GFM product is a Copernicus

operational service that provides near-real-time global flood monitoring based on incoming Sentinel-1 SAR acquisitions.

The detailed revisions in the text are as follows:

The Sentinel-1 Global Flood Monitoring product is an operational near-real-time flood mapping service of the Copernicus Emergency Management Service. It processes incoming Sentinel-1 SAR acquisitions globally and derives flood extent using an ensemble of three independent automated flood-mapping algorithms. The final observed flood extent is generated by a majority-decision ensemble and is accompanied by reference water, exclusion mask, likelihood, and advisory-flag layers. Although GFM provides global near-real-time flood information and a historical Sentinel-1 archive from 2015 onward, its effective availability and reliability depend on the Sentinel-1 acquisition schedule and on exclusion-mask conditions such as topographic distortion, radar shadow, dense vegetation, urban areas, and low-sensitivity surfaces. Therefore, for the urban flooding extraction in Weihui City, the Sentinel-1 Global Flood Monitoring product appears to be ineffective. The accuracy of flood extraction from the natural areas of this product is slightly lower than that of the method used in this study.

Comment 3: the authors should also compare it to the simplest SAR method, such as Otsu or a simple image differencing. How much better is it to justify the increased computational complexity?

Response:

We thank the reviewer for this suggestion. We added comparisons with Otsu thresholding and simple SAR image differencing. These methods were implemented using the same Sentinel-1 acquisitions and validation masks to ensure a fair comparison.

The results show that Otsu and simple differencing can capture large open-water flood regions but are less stable in heterogeneous urban and vegetated areas. Otsu thresholding was particularly sensitive to the absence of a clear bimodal distribution in urban Z-score images, while simple differencing was affected by non-flood backscatter changes caused by vegetation growth, rainfall, and surface roughness.

We also clarified that the additional computational complexity of the proposed method is moderate because the main operations are temporal mean/std calculation, Z-score transformation, and rule-based classification. These steps can be implemented efficiently for large areas.

We added baseline comparisons with Otsu and image differencing in Section 4 and expanded the Discussion to justify the computational cost.

Response to Reviewer3 (CC-community)

Dear Community,

We sincerely thank the reviewer for the positive evaluation of our manuscript. We appreciate the constructive comments, which helped us improve the clarity, methodological justification, validation discussion, and presentation of the manuscript. Our point-by-point responses are provided below.

Comment 1: The manuscript would benefit from careful English editing, as there are many grammatical issues and repeated singular/plural mismatches such as “images is” and awkward phrasing throughout the text.

Response:

Thank you for pointing this out. We have carefully revised the language throughout the manuscript. In particular, we corrected grammatical errors, singular/plural mismatches, awkward expressions, and inconsistent sentence structures. We also improved the readability of the Abstract, Introduction, Methods, Results and Discussion, and Conclusions. Expressions such as “images is” have been corrected, and several long or unclear sentences have been rewritten for clarity.

Comment 2: The novelty relative to prior Sentinel-1 time-series flood mapping studies, especially Z-score or anomaly-based approaches, should be stated more explicitly in the Introduction and Discussion. At present, the distinction from earlier methods is somewhat diffuse.

Response:

Thank you for this important suggestion. We have strengthened the description of the novelty in the Introduction and Discussion. The revised manuscript now explicitly distinguishes our method from previous Sentinel-1 flood mapping studies by emphasizing three aspects: first, pixel-wise temporal standardization is used to reduce dependence on event-specific global thresholds; second, VV and VH anomalies are integrated in a rule-based classification tree to handle different flood responses in natural and built-up areas; third, the method remains operationally scalable because it only requires Sentinel-1 GRD time series, auxiliary water-occurrence data, and simple statistical operations. These points clarify how our method differs from conventional single-image thresholding, image differencing, and single-polarization anomaly detection. The revised Introduction now explicitly states these three contributions.

Revision made:

We added a paragraph in the Introduction to explicitly summarize the methodological contributions:

1. pixel-wise temporal Z-score standardization;
2. dual-polarization VV/VH classification-tree integration;
3. operational scalability for both large-area flood monitoring and urban inundation detection.

We also expanded the Discussion by comparing our method with VV-only Z-score, VH-only Z-score, simple image differencing, Otsu thresholding, and the Sentinel-1 Global Flood Monitoring product. The revised comparison shows that the proposed method has more balanced performance across both natural and urban environments.

Comment 3: The choice of thresholds for natural areas (-2, -2.5, -3) and urban areas (2, 2.5, 3) is statistically motivated, but the final recommended threshold should be justified more systematically, perhaps with a table summarizing OA, UA, PA, and CSI for all cases.

Response:

Thank you for this helpful suggestion. We have revised the threshold selection discussion to provide a clearer statistical and empirical justification. The revised manuscript explains that negative Z-score thresholds are used for natural areas because open-water flooding generally decreases SAR backscatter due to specular reflection, whereas positive Z-score thresholds are used for urban areas because inundation around buildings can enhance wall–water double-bounce scattering and increase backscatter intensity. We also clarified that VV and VH are standardized independently, so the same numerical Z-score threshold represents the same standardized anomaly magnitude within each polarization rather than the same absolute backscatter change in dB.

Revision made:

We added a clearer explanation of threshold interpretation in the Methods section and expanded the Results section to discuss why -2 was selected for natural areas and 2 for urban areas. In natural areas, lowering the threshold captures more flood pixels but introduces more speckled false alarms; at the 95% confidence level, the method achieved approximately 90% OA and CSI above 60%, supporting the use of -2. In urban areas, lower thresholds improved flood-region continuity and increased CSI, although UA slightly declined.

We also added a comparison table among different methods. The current Table 2 reports CSI/OA for the proposed method and baseline methods: the proposed dual-polarization Z-score tree achieved 60%/90% CSI/OA in natural areas and 62%/73% in building areas.

Comment 4: The validation strategy deserves clearer discussion, especially because the Sentinel-2 and Sentinel-1 acquisitions are not fully synchronous, and the paper itself notes that flood extent changed rapidly between dates. This temporal mismatch should be treated more explicitly as a source of uncertainty.

Response:

Thank you for this important comment. We agree that temporal mismatch between Sentinel-1 and Sentinel-2 may introduce uncertainty, especially during rapidly changing flood events. We have revised the validation section to explicitly discuss this issue. The manuscript now states that the time difference between Sentinel-2 and Sentinel-1 was approximately 8–9 hours, while the UAV imagery used for urban validation had an approximately 2-hour time difference. We also clarified that the UAV validation should be regarded as local high-resolution validation rather than complete city-wide validation because of its limited spatial coverage.

Comment 5: Please define all abbreviations at first use and keep terminology consistent, for example “building areas,” “urban areas,” and “urban (building) areas,” which are currently used interchangeably.

Response:

Thank you for this comment. We have checked the manuscript and revised the abbreviation definitions

and terminology. Abbreviations such as SAR, Sentinel-1 GRD, GPM, IMERG, GSW, DEM, UAV, MNDWI, OA, UA, PA, and CSI have been defined at first use. We also revised the wording related to urban inundation areas. To improve consistency, we now use “building areas” consistently when referring to built-up regions where urban inundation is detected through enhanced double-bounce scattering. We use “natural areas” for non-artificial surfaces such as water bodies, bare soil, cropland, and forests.

Comment 6: In the methods section, it would help to clarify how the “reference images” were selected, how many were used for each flood-date analysis, and whether seasonality was controlled equally for natural and urban surfaces.

Response:

Thank you for this suggestion. We agree that the reference-image selection procedure should be described more clearly. We have revised the Methods section to clarify that the Z-score calculation was based on same-orbit Sentinel-1 time series, so that differences caused by orbit direction, polarization, and imaging geometry were minimized. We also clarified that the baseline statistics were calculated separately for each polarization and orbit configuration.

Revision made:

We added text explaining that same-orbit reference images were used to calculate the temporal mean and standard deviation for each pixel, and that flood-period images were standardized relative to this reference distribution. For natural areas, seasonality is important because vegetation phenology can strongly affect backscatter stability; therefore, reference images were selected with attention to seasonal consistency. For urban/building areas, the manuscript now explains that building backscatter was comparatively stable during non-flood periods and was less affected by seasonal vegetation dynamics, so seasonal constraints are less strict than in cropland or other vegetation-covered natural areas. The manuscript also added normality diagnostics for different land-cover classes, showing that urban VH backscatter was relatively stable while cropland had stronger deviations due to crop phenology.

Comment 7: Figures are generally helpful, but some captions are too brief. The reader would benefit from more explicit legends, clearer color explanations, and a more direct connection between figure content and the interpretation in the main text, especially for Figs. 5–8.

Response:

Thank you for this helpful suggestion. We have revised the figure captions and improved the discussion associated with Figs. 5–8. We added clearer descriptions of what each color represents, the date of each image, and how the figure supports the interpretation in the main text.

Revision made:

For Fig. 5 and Fig. 6, we clarified that the panels compare flood monitoring results from different dates and that differences among these panels reflect the rapid temporal evolution of flood extent. For Fig. 7, we clarified that grey indicates areas not affected by flood and red indicates areas affected by flood, and that the UAV imaging range is shown in Fig. 1. For Fig. 8, we clarified that the UAV oblique images represent severely inundated and moderately inundated urban regions, and we added an explanation of how low-rise building height and water depth can weaken or suppress wall–water

double-bounce scattering. The revised text now directly links the UAV visual evidence to SAR missed-detection mechanisms.

Comment 8: I strongly recommend that the authors cite the recent paper *Spatiotemporal SWMM-LSTM surrogate modeling for efficient node-level water depth and inflow prediction in urban drainage networks* as it is highly relevant to the topic of urban flood prediction and would strengthen the manuscript's engagement with recent advances in the field.

Response:

Thank you for recommending this relevant recent study. We agree that this paper is closely related to urban flood prediction and data-driven urban drainage modeling. We have added it to the Introduction/Discussion to broaden the manuscript's engagement with recent advances in urban flood modeling. The paper is authored by Nima Zafarmomen, Vidya Samadi, and Edoardo Borgomeo, published in Cambridge Prisms: Water in 2026, with DOI 10.1017/wat.2026.10013.