Responses to Reviewers

Please note that all line numbers cited in the responses correspond to the revised manuscript, which is not submitted simultaneously at this stage.

Table des matières

1 Response to Reviewer #1 2

1.1 Comment 1 2

1.2 Comment 2 3

1.3 Comment 3 4

2 Response to Reviewer #2 6

2.1 Comment 1 6

2.2 Comment 2 8

2.3 Comment 3 12

2.4 Comment 4 13

2.5 Comment 5 14

2.6 Comment 6 16

2.7 Comment 7 18

3 Response to Reviewer #3 20

3.1 Comment 1 20

3.2 Comment 2 21

3.3 Comment 3 21

3.4 Comment 4 22

3.5 Comment 5 24

1 Response to Reviewer #1

We thank the reviewer for the careful reading of our manuscript and for the positive assessment of the study. Our detailed responses to each comment are provided below.

1.1 Comment 1

Lines 74-82: The discussions on the performance of SWOT from various sources need to be summarized in a coherent manner. There are three scales involved: a global assessment of 0.15 m (Yu et al,2024); a regional assessment of 0.25 m (Jiang et al 2025), and a subcontinental (over India) assessment of 18 cm (Patidar et al 2025). The latter two are for narrow rivers, but the first seems to be for all rivers. I think these results need to be discussed in the context of spatial scales versus the mission requirement.

Response:

We agree with the reviewer that the discussion of SWOT performance across different studies required clearer organization and explicit framing in terms of spatial scale and mission requirements, and the manuscript was revised accordingly. A new synthesized paragraph has been added between Lines 75-107, which now reads as follows:

"Subsequent validation studies have shown that the reported accuracy of SWOT WSE measurements varies with the study region, river width, and product type (PIXC, RiverSP node, and RiverSP reach), with errors decreasing because of spatial averaging.

At the global scale, based on a worldwide evaluation including rivers of diverse width and hydrological conditions and using satellite-based WSE reference datasets such as Hydroweb (https://www.theia-land.fr/blog/product/hauteur-des-lacs-et-rivieres/) and G-REALM (https://ipad.fas.usda.gov/cropexplorer/global_reservoir/), the average measurement error is estimated to be below 0.15 m (Yu et al., 2024), which is consistent with the expected order of magnitude of the mission's performance. At the subcontinental scale, validation over India based on more than 14,000 observations across 419 hydrometrics stations showed relative errors of about 0.18 m for rivers wider than 100 m and about 0.26 m for narrower rivers, based on node-level evaluations, highlighting a clear dependence of SWOT WSE accuracy on river width (Patidar et al., 2025). These results consistently indicate that node-level SWOT WSE accuracy decreases for narrower rivers, while remaining within the range of typical in situ measurement uncertainties. At finer spatial scales, a watershed-scale assessment evaluated satellite performance using node-level WSE data. The SWOT Science Requirements are defined for reach-scale products, whereas the present assessment relies on node-level observations. Comparison with the mission requirements is therefore not appropriate. During the Cal/Val orbit, validation against United States Geological Survey (USGS) gauging records, based on 39 stations, showed a median unbiased Root Mean Square Error (ubRMSE) of about 0.25 m, even for rivers narrower than 100 m and in some cases below 50 m (Jiang et al., 2025). This level *of agreement is consistent with typical uncertainties in ground-based WSE measurements used in flood applications in the United States, which are commonly on the order of 0.15 to 0.30 m according to USGS guidelines (NOAA, 2011). In contrast to global-scale evaluations based on reach-averaged products, the present analysis relies on node-level WSEs, for which error metrics are expected to be larger due to the absence of spatial averaging. Nevertheless, studies specifically addressing SWOT performance in narrow rivers remain limited, particularly during a flood event. Moreover, most existing assessments rely on earlier versions of the SWOT products (Version C), and primarily assess relative errors, in which potential systematic biases between SWOT WSE and reference observations are removed, while a more recent Version D is now available.*

In this context, the present study investigates the ability of SWOT observations to support the analysis of hydraulic consistency and the calibration and validation of hydraulic models in a narrow river, using the most recent RiverSP Version D product. The analysis focuses on the Du Gouffre River (Quebec, Canada; ≈ 40 m wide), which experienced a major flood in May 2023 with an estimated return period of 60 years, based on three-hour averaged flows (COMEXI-RDG, 2023), and for which a LISFLOOD-FP model (Bates and De Roo, 2000) was available. This rare and extreme event offers a valuable opportunity to assess SWOT performance under flood conditions in a narrow river through a concrete hydraulic case study. The Du Gouffre River sector was covered by the SWOT CAL/VAL orbit, enabling daily observations of the flood event. Unlike most previous studies, the evaluation is based on the smoothed WSE variable (wse_sm) from Version D and relies on a direct comparison of absolute WSEs, ensured through a consistent geoid conversion between SWOT observations and model outputs."

1.2 Comment 2

Lines 219-223: The discussion here is very confusing. On Fig 4, I don't find the information on probability, the 68th percentile is 0.05 m, RMSE 0f 0.24 m? Something is missing on Fig 4, which shows only the comparison of WSE from SWOT with model simulations.

Response:

We thank the reviewer for pointing out the confusion in Lines 219-223. We agree that the discussion referring to probabilities, percentiles, and RMSE was not clearly supported by the originally referenced figure.

This confusion arose because the discussion was mistakenly associated with Fig. 4, which indeed shows only the comparison between SWOT WSEs and LISFLOOD-FP simulations. To address this issue, we have added a new figure (now **Figure 3**) that explicitly presents the distribution of absolute SWOT WSE errors with respect to in situ observations at the Saint-Joseph tide gauge. This figure includes (a) the likelihood of occurrence expressed as a cumulative distribution function (CDF) and (b) a boxplot summarizing the statistical distribution of the absolute errors.

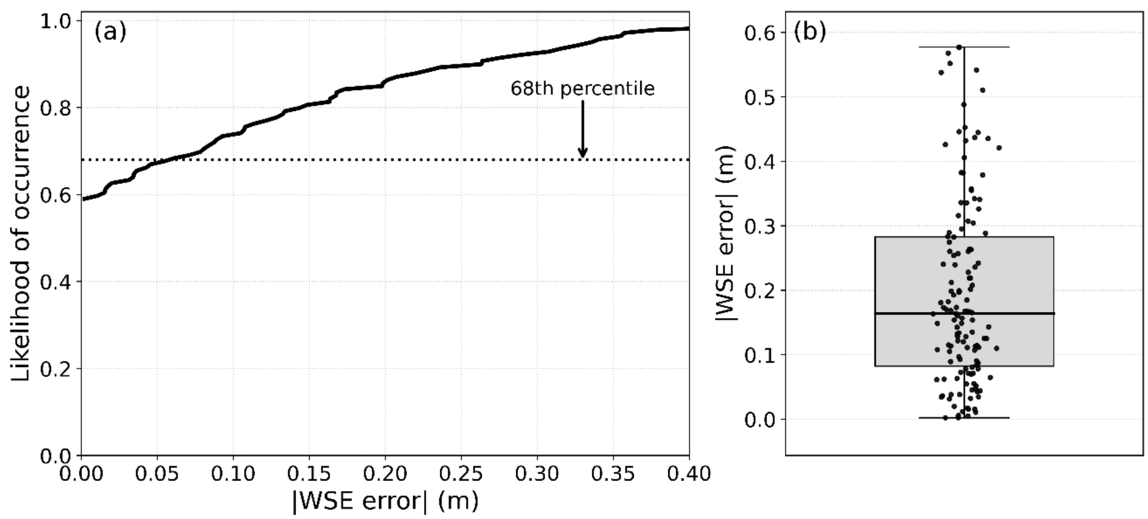

Figure 3: (a) Likelihood of occurrence of absolute WSE errors between SWOT observations and measurements at the Saint-Joseph-de-la-Rive tide gauge station. The black curve represents the empirical cumulative distribution function of the absolute WSE differences. The horizontal dotted line indicates the 68th percentile of the distribution. (b) The boxplot illustrates the dispersion of absolute WSE errors.

Lines 258-265 *"Figure 3a shows the probability that the absolute WSE error is less than or equal to a given value (Likelihood of occurrence), indicating that the 68th percentile (corresponding to the 1-sigma metric defined in the SWOT requirements) is 0.05 m, with an RMSE of 0.24 m. The boxplot (Figure 3b) complements this probabilistic analysis by illustrating the distribution of absolute WSE errors (m). It shows a low median and a relatively limited interquartile range, indicating that the errors are well centred and not dominated by systematic bias or by extreme deviations in the error distribution. The SWOT observations used are of good quality (node_q < 3 et xovr_cal_q < 2). The obtained values meet the accuracy thresholds defined by the SWOT mission, which corresponds to the expected accuracy of 0.10 m averaged over 1 km² or 0.45 m averaged over 0.01 km² (Desai, 2018)."*

1.3 Comment 3

Lastly, I would suggest that all the figure captions be more elaborate with some details.

Response:

We agree with this suggestion and have revised all figure captions accordingly. Each caption now provides sufficient detail to allow the figure to be understood independently of the main text. We believe that these revisions significantly improve the readability and clarity of the figures.

1. Figure 1: Du Gouffre River study area

Figure 1: Study area and spatial context of the Du Gouffre River in Québec, Canada. (A) Location of the Du Gouffre watershed and its main tributaries (Gros Bras River, Des Mares River, and Bras du Nord-Ouest River), together with the SWOT calibration orbit coverage. The positions of the hydrometric station Saint-Urbain (051305), the tide station Saint-Joseph-de-la-Rive (03057), and the La-Galette station are indicated. The inset map shows the location of the study area within the province of Québec. (B) Detailed view of the river reaches analyzed in this study near Baie-Saint-Paul, showing the SWOT river reaches and the LISFLOOD-FP model domain.

2. Figure 2: Instantaneous discharge (black curve) measured at hydrometric station 051305, and the WSE (blue curve) recorded at the tide gauge station 03057 (Station 03057, 2025), from April 25 to May 7, 2023. The black dots indicate the dates of SWOT satellite acquisitions during its calibration orbit. The peak of the flood (reaching nearly 500 m^3/s) occurred on May 1, 2023, at 12:30 p.m. (Local Time).

Figure 2: Instantaneous discharge (black curve) measured at hydrometric station 051305, and the WSE (blue curve) recorded at the tide gauge station 03057, from April 25 to May 7, 2023. Vertical dashed lines indicate the dates of SWOT satellite acquisitions during the calibration orbit, with cycle numbers labeled. The flood peak occurred on 1 May 2023 at approximately 12:30 local time, reaching a discharge close to 500 $m^3 s^{-1}$.

Figure 3: (a) Likelihood of occurrence of absolute WSE errors between SWOT observations and measurements at the Saint-Joseph-de-la-Rive tide gauge station. The black curve represents the empirical cumulative distribution function of the absolute WSE differences. The horizontal dotted line indicates the 68th percentile of the distribution. (b) The boxplot illustrates the dispersion of absolute WSE errors.

3. Figure 3: Comparison of SWOT and LISFLOOD-FP WSEs during the flood peak (cycle 508, May 1, 2023)

Figure 4: Longitudinal profile of WSE along the Du Gouffre River during the flood peak observed by SWOT (cycle 508, 1 May 2023). Black dots represent SWOT node-level WSE observations. The blue and orange lines show WSE simulated by the LISFLOOD-FP model using the calibrated and non-calibrated configurations, respectively. The black line corresponds to the riverbed elevation. The locations of the Bras du Nord-Ouest and Des Mares tributaries are indicated by colored symbols. Vertical dotted lines denote the limits between SWOT river reaches, and the location of a bridge structure is indicated.

Figure 4: Comparison of SWOT and LISFLOOD-FP WSEs for all cycles before and after the May 1 flood (excluding Cycle 508); the maximum and minimum discharges based on the uncertainty of the rating curve are also represented.

Figure 5: Longitudinal profiles of WSE along the Du Gouffre River for all SWOT overpasses before and after the 1 May 2023 flood peak, excluding cycle 508. For each cycle, SWOT node-level WSE observations (black dots) are compared with LISFLOOD-

FP simulated WSEs, including the profile (blue line) and the envelope defined by the minimum and maximum discharges derived from rating-curve uncertainty (shaded area). The riverbed elevation is shown by the black line. The locations of the Bras du Nord-Ouest and Des Mares tributaries, the bridge structure, and the limits between SWOT reaches are indicated.

4. Figure 5: Cumulative distribution of SWOT quality indicators (wse_u in meters, dark_frac unitless) across cycles.

Figure 6: Likelihood of occurrence of SWOT node-level quality indicators for the Du Gouffre River: (A) wse_u (m) and (B) dark_frac (–). The flood-peak acquisition (cycle 508) is compared with all remaining SWOT cycles excluding cycle 508.

2 Response to Reviewer #2

We thank the reviewer for taking the time to review the manuscript. Our responses to the reviewer's comments are provided below.

2.1 Comment 1

While study is a good case study, its scientific novelty and strength of conclusions remain limited. The manuscript demonstrates consistency between SWOT WSE and a pre-existing LISFLOOD-FP model, rather than providing a rigorous, independent assessment of SWOT performance in narrow rivers. As a result, contribution is not significant and shows simply as case study rather than advancing understanding beyond what has already been shown in recent studies, especially for the narrow river (<50 m).

Response:

We acknowledge that the manuscript could have more clearly framed the results as those of a case study rather than implying broader conclusions for narrow rivers in general. We agree that the present work does not aim to provide a generalized or fully independent assessment of SWOT performance across all rivers narrower than 50 m.

To date, most studies focusing on SWOT observations in narrow rivers primarily assess measurement performance through relative comparisons with in situ data or other satellite products, often with the objective of evaluating SWOT accuracy. In contrast, practical applications of SWOT-derived WSEs for hydraulic modelling remain very limited, particularly for small rivers and under real flood conditions. In this study, SWOT observations are converted into a consistent vertical datum and directly integrated within a hydraulic modelling framework, which constitutes one of the first demonstrations of the operational use of SWOT data for hydraulic analysis in a narrow river context.

The originality of the present work therefore lies not in re-evaluating SWOT measurement accuracy, but in demonstrating the usefulness of SWOT-derived WSEs for understanding and analysing an extreme flood event in a poorly gauged narrow river. By exploiting SWOT observations acquired during the calibration/validation (CAL/VAL) orbit, the study documents the behaviour of WSEs across different phases of the flood hydrograph, including pre-event, peak-flood, and post-event conditions, and illustrates how these data can support the interpretation of discharge dynamics and hydraulic consistency during an extreme event.

The manuscript has been revised thoroughly to more explicitly position the study as a site-specific case study and to clarify that the conclusions are limited to the investigated river and flood event. Rather than aiming to generalize SWOT performance for all narrow rivers, this work demonstrates the added value of SWOT observations as a complementary source of information for hydraulic studies, particularly in data-scarce environments and under extreme hydrological conditions. We believe that this contribution helps pave the way for future applications of SWOT data in flood monitoring and hydraulic modelling of narrow rivers.

*"**Abstract.** Floods are among the most frequent and damaging natural hazards worldwide, and reliable observations of water surface elevation (WSE) are essential for improving flood modelling and risk management. The Surface Water and Ocean Topography (SWOT) satellite, launched in 2022, offers new opportunities to monitor river hydrodynamics from space, but its performance in relatively narrow rivers (< 50 m width) remains poorly documented. This study evaluates the potential of SWOT WSEs for flood monitoring **through a site-specific hydraulic application** by comparing them with in situ observations as well as simulations from an existing large-scale hydraulic model (LISFLOOD-FP) on the Du Gouffre River (width ≈ 40 m), located in Quebec, Canada. The L2_HR_RiverSP (RiverSP) SWOT product Version D, derived from a priori database (SWORD-version 17b), was first compared with one-minute WSE measurements from a tidal gauge located downstream the Du Gouffre River in the St. Lawrence River. **This comparison, based on in situ reference measurements, confirmed the overall quality of the SWOT data in this area, with a Root Mean Square Error (RMSE) of 0.24 m.** Then, a major flood event (with a return period of about 60 years) which occurred on May 1, 2023, during the SWOT's calibration orbit, was used to conduct a daily analysis of the entire flood event. Eleven observation cycles, covering the period from April 25 to May 7, 2023, were analysed. Limited ground-based observations were available along the studied reach during the flood, **highlighting the added value of SWOT observations in this data-scarce context**. The 1D/2D hydraulic model LISFLOOD-FP was run for the discharges corresponding to eleven SWOT cycles. **Comparisons between SWOT-derived WSEs and model-simulated WSEs yielded biases ranging from -0.30 to 0.43 m and RMSE values between 0.22 m and 0.54 m, indicating generally, consistent behavior between observed and simulated WSEs across the analyzed cycles. During the flood peak on 1 May, larger discrepancies were observed, reflecting uncertainties in upstream discharge estimates under extreme***

flow conditions. In this context, SWOT-derived WSEs provided complementary information that helped diagnose discharge underestimation during the peak event. These results are not intended as an independent validation of SWOT measurement accuracy but are specific to the studied river and flood event and demonstrate the practical usefulness of SWOT observations for hydraulic studies in narrow rivers. They highlight the potential contribution of SWOT data for flood monitoring and hydraulic modelling in similar data-scarce contexts, particularly under extreme hydrological conditions."

Introduction:

Lines 99-107 *"In this context, the present study investigates the ability of SWOT observations to support the analysis of hydraulic consistency and the calibration and validation of hydraulic models in a narrow river, using the most recent RiverSP Version D product. The analysis focuses on the Du Gouffre River (Quebec, Canada; ≈ 40 m wide), which experienced a major flood in May 2023 with an estimated return period of 60 years, based on three-hour averaged flows (COMEXI-RDG, 2023), and for which a LISFLOOD-FP model (Bates and De Roo, 2000) was available. This rare and extreme flood event offers a valuable opportunity to assess SWOT performance under flood conditions in a narrow river through a concrete hydraulic case study. The Du Gouffre River sector was covered by the SWOT CAL/VAL orbit, enabling daily observations of the flood event. Unlike most previous studies, the evaluation is based on the smoothed WSE variable (wse_sm) from Version D and relies on a direct comparison of absolute WSEs, ensured through a consistent geoid conversion between SWOT observations and model outputs."*

Discussion and Conclusion:

Lines 338-342 *"Within this constrained observational context, the present study is therefore framed as a site-specific case study that explores the practical use of SWOT observations to support hydraulic analysis during a rare and extreme flood event, rather than as a strict assessment of SWOT measurement accuracy. The following discussion focuses on how SWOT-derived WSEs can complement sparse in situ data and contribute to the evaluation of hydraulic consistency and discharge dynamics in a narrow river under extreme flow conditions."*

Lines 383-358 *"Rather, within the context of this site-specific case study, it illustrates how SWOT observations can provide complementary information to diagnose discharge underestimation in a poorly gauged tributary and to enhance the internal consistency of hydraulic model inputs during an extreme flood event."*

2.2 Comment 2

It looks like the manuscript is the lack of independence between the SWOT observations and data and LISFLOOD-FP model used for evaluation. The LISFLOOD-FP model is *based on LiDAR-derived bathymetry and unchanging roughness assumptions, and discharge inputs that are uncertain during the flood peak, and most important issue this model is not calibrated. However, SWOT WSEs are then compared against this model and refereed "accurate" whenever good agreement is found. Additionally, during the flood peak (cycle 508), SWOT is explicitly used to calibrate tributary discharge, after which agreement improves. This creates issue, as SWOT validates a model that is partially constrained or adjusted using SWOT itself. Consequently, the reported RMSE and bias values cannot be interpreted as true SWOT measurement errors.*

Response:

(1) Independence between SWOT observations and LISFLOOD-FP

The concern regarding a potential lack of independence between SWOT observations and the LISFLOOD-FP model is understandable. In this study, SWOT and LISFLOOD-FP are treated as independent data sources for all analyzed cycles before and after the flood peak. The only exception is the peak-flood cycle (cycle 508), which is explicitly identified and treated separately in the analysis, as SWOT-derived WSEs were used during this cycle to adjust underestimated tributary inflows when standard discharge transposition methods proved insufficient.

(2) LISFLOOD-FP model calibration

While Manning's roughness coefficient is kept fixed in the LISFLOOD-FP setup, the model should not be considered uncalibrated. In the absence of direct bathymetric surveys, as is often the case in large-scale or data-scarce river systems, calibration strategies may rely on adjustments of channel bed elevation rather than friction parameters. In this study, the hydraulic model incorporates a LiDAR-derived bathymetry that was previously adjusted using an inverse hydraulic approach based on known discharge conditions at the time of LiDAR acquisition (Choné et al., 2021, 2024). This adjustment constitutes a form of model calibration, even though no calibration of Manning's roughness coefficient was performed, in contrast to many hydraulic modelling studies. This modelling strategy has been clarified in the revised manuscript to avoid ambiguity regarding the calibration status of the model. Outside the flood peak, LISFLOOD-FP simulations are in good agreement with upstream in situ WSE at the Saint-Urbain hydrometric station, as reflected by small residual errors. This agreement provides additional confidence in the model's performance.

Lines 160-163 *"The model is fed solely by remote-sensed data, incorporating an inverse hydraulic model to estimate bed elevation from LiDAR water surfaces, using the known discharge value on the LiDAR day of acquisition (Choné et al., 2021, 2024).* ***While a constant Manning's roughness coefficient is prescribed over the study area, the adjustment of bed elevation constitutes the main calibration step of the hydraulic model."***

Lines 329-337 *"The interpretation of the results must be placed within the context of limited ground-based observations available along the studied reach. Aside from a downstream tidal gauge and a single upstream hydrometric station, no spatially distributed in situ WSE measurements were available during the flood event. Despite this limitation, preliminary comparisons provide useful independent benchmarks that help contextualize the analysis. First, the comparison between SWOT-derived WSEs and one-minute observations at the downstream tidal gauge shows good agreement, with a 68th percentile absolute error of 0.05 m and an RMSE of 0.24 m (Fig. 3), consistent with the SWOT mission accuracy requirements.* ***In addition, an upstream comparison between LISFLOOD-FP simulations and in situ WSE observations at the Saint-Urbain hydrometric station yield a bias of 0.07 m and an RMSE of 0.12 m for all cycles except the flood peak, supporting the overall consistency of the hydraulic model outside peak-flow conditions.****"*

(3) Interpretation of agreement between SWOT and LISFLOOD-FP

We agree with the reviewer that referring to SWOT WSEs as "accurate" based on agreement with LISFLOOD-FP can be misleading. Accordingly, the manuscript has been revised to consistently frame these comparisons as an assessment of model-SWOT consistency, rather than as an independent validation of SWOT measurement accuracy. This applies to all analyzed cycles, not only to the flood peak.

Lines 249-252 *"To compare WSE from SWOT observations and hydraulic modelling, two statistical indicators were calculated for each overpass date: Root Mean Square Error (RMSE), used to assess consistency between the values produced by the LISFLOOD-FP hydraulic model and those measured by the SWOT satellite, both subject to uncertainty, and the bias, which highlights any systematic trends of overestimation or underestimation."*

Lines 294-302 *"The observed flood is covered by cycles 508, 509, and 510, with cycle 508 corresponding to the flood peak on 1 May. For cycle 508, a comparison between SWOT-derived WSEs and LISFLOOD-FP simulations prior to any adjustment of the Des Mares tributary inflow shows a bias of 1.30 m and an RMSE of -1.21 m. To investigate the origin of this discrepancy, a discharge adjustment was performed for the Des Mares tributary using SWOT-derived WSEs. Increasing the tributary inflow to 240 m³/s results in a bias of -0.30 m and an RMSE of 0.54 m (Fig. 4). The comparison before and after the adjustment highlights an improvement in model-SWOT consistency during the flood peak. Because SWOT observations are used to inform the tributary discharge adjustment for this cycle, this comparison is not interpreted as an independent validation of SWOT measurement accuracy, but rather as an assessment of consistency between the adjusted hydraulic simulation and SWOT-derived WSEs.*"

Lines 311-313 *"Unlike cycle 508, the discharges for this overpass were well constrained and did not involve significant tributary inputs, which likely contributed to the good consistency between the two datasets."*

Lines 343-344 *"The results indicate a general consistency between SWOT observations and the WSEs simulated by the LISFLOOD-FP model across most of the analyzed cycles."*

Lines 378-388 *"During the flood peak on 1 May (cycle 508), the initial comparison between SWOT-derived WSEs and LISFLOOD-FP simulations showed a pronounced discrepancy. After adjusting the inflow of the Des Mares tributary, the agreement improved, yielding a bias of -0.30 m and an RMSE of 0.54 m. Compared with the pre-adjustment simulations, this result indicates a improved consistency between the simulated WSEs and the SWOT observations during peak-flow conditions. As the tributary inflow is adjusted using SWOT-derived WSEs for this cycle, the resulting improvement should not be interpreted as an independent assessment of SWOT measurement accuracy. Rather, within the context of this site-specific case study, it illustrates how SWOT observations can provide complementary information to diagnose discharge underestimation in a poorly gauged tributary and to enhance the internal consistency of hydraulic model inputs during an extreme flood event. Similar conclusions were drawn by Diouf et al. (2025), who showed that SWOT observations (using the L2_HR_PIXC product) can support hydraulic model calibration in poorly instrumented environments by constraining hydraulic parameters and improving the representation of WSE dynamics in areas lacking in situ measurements."*

(4) Use of SWOT during the flood peak (cycle 508)

We agree that, for cycle 508, RMSE and bias values cannot be interpreted as true SWOT measurement errors, as SWOT observations are used to inform the adjustment of tributary discharge during this specific period. This step was undertaken only after identifying a non-systematic but persistent bias between SWOT and LISFLOOD-FP across pre- and post-event cycles, together with strong temporal and spatial coherence between the two datasets. Additional confidence in SWOT WSE quality in this area is provided by independent comparisons with downstream tide-gauge observations. In a context of limited hydrometric data availability, this ensemble of evidence motivated the exploratory use of SWOT observations to correct an underestimation of discharge during the extreme flood peak. This adjustment is now clearly isolated in the manuscript and is not presented as a validation of SWOT accuracy.

Lines 30-32 *"These results are not intended as an independent validation of SWOT measurement accuracy but are specific to the studied river and flood event and demonstrate the practical usefulness of SWOT observations for hydraulic studies in narrow rivers. "*

Lines 338-342 *"Within this constrained observational context, the present study is therefore framed as a site-specific case study that explores the practical use of SWOT* *observations to support hydraulic analysis during a rare and extreme flood event, rather than as a strict assessment of SWOT measurement accuracy. The following discussion focuses on how SWOT-derived WSEs can complement sparse in situ data and contribute to the evaluation of hydraulic consistency and discharge dynamics in a narrow river under extreme flow conditions."*

2.3 Comment 3

The absence of independent, spatially distributed in situ water level data along the studied river reach further lacks sufficient validation. Apart from the downstream tide gauge, no ground-based WSE measurements are available during the flood event, especially along the main channel where SWOT nodes are evaluated. As a result, the performance of SWOT upstream is referred entirely from agreement with the hydraulic model rather than from direct observations. While drone imagery provides useful qualitative confirmation of flood extent, it does not offer independent vertical validation of WSE. This limitation is referred only briefly but should be emphasized more strongly, as it fundamentally constrains the strength of the conclusions regarding SWOT accuracy in narrow rivers.

Response:

We agree with the reviewer that the absence of spatially distributed in situ WSE measurements along the studied river reach represents a key limitation of this study.

In response to this comment, the manuscript has been revised to more explicitly emphasize this limitation and to further temper the interpretation of the results. In particular, the conclusions have been reframed to avoid claims of independent accuracy assessment and to clearly state that the analysis primarily demonstrates model-SWOT consistency under extreme flood conditions, rather than direct validation against spatially distributed in situ observations.

This constraint is now highlighted more prominently, and the study is positioned as a case study documenting SWOT behavior during a rare flood event in a narrow river.

Lines 329-342 *"The interpretation of the results must be placed within the context of limited ground-based observations available along the studied reach. Aside from a downstream tidal gauge and a single upstream hydrometric station, no spatially distributed in situ WSE measurements were available during the flood event. Despite this limitation, preliminary comparisons provide useful independent benchmarks that help contextualize the analysis. First, the comparison between SWOT-derived WSEs and one-minute observations at the downstream tidal gauge shows good agreement, with a 68th percentile absolute error of 0.05 m and an RMSE of 0.24 m (Fig. 3), consistent with the SWOT mission accuracy requirements. In addition, an upstream comparison between LISFLOOD-FP simulations and in situ WSE observations at the Saint-Urbain hydrometric station yield a* *bias of 0.07 m and an RMSE of 0.12 m for all cycles except the flood peak, supporting the overall consistency of the hydraulic model outside peak-flow conditions.*

Within this constrained observational context, the present study is therefore framed as a site-specific case study that explores the practical use of SWOT observations to support hydraulic analysis during a rare and extreme flood event, rather than as a strict assessment of SWOT measurement accuracy. The following discussion focuses on how SWOT-derived WSEs can complement sparse in situ data and contribute to the evaluation of hydraulic consistency and discharge dynamics in a narrow river under extreme flow conditions."

2.4 Comment 4

The discharge calibration for the Des Mares tributary during the flood peak is presented as a key demonstration of SWOT's advantage, but I feel it is not sufficiently justified/analyzed to support these strong claims. Authors presented a single calibrated scenario and fail to provide any sensitivity analysis to demonstrate that the improved agreement is exclusively attributable to the adjusted tributary discharge. Other factors, such as roughness values, downstream boundary conditions, or bathymetric uncertainty, could possibly produce similar improvements. As such, the conclusion that SWOT observations reveal discharge underestimation and effectively constrain ungauged tributary inflows is reasonable but not conclusively demonstrated.

Response:

Sensitivity analyses were performed on Manning's roughness coefficient and on the downstream boundary condition and have been added to the manuscript. These tests showed that reasonable variations in roughness values and downstream WSE did not significantly reduce the discrepancy observed during the flood peak and were therefore insufficient to improve the agreement between SWOT-derived WSEs and the model simulations. This indicates that the tributary discharge adjustment was the main factor controlling model-SWOT consistency during cycle 508.

Lines 187-195 *"In addition, sensitivity analyses were performed to assess the influence of key modelling assumptions on the simulated WSEs during the flood peak. In particular, the Manning's roughness coefficient and the downstream boundary condition were varied within physically plausible ranges. These tests showed that such variations had a limited impact on the simulated WSEs and did not significantly reduce the discrepancies observed between LISFLOOD-FP simulations and SWOT-derived WSEs during cycle 508. Moreover, parameter adjustments that slightly improved agreement during the flood peak systematically led to a deterioration of model-SWOT consistency for the other cycles. This indicates that uncertainties related to channel roughness or downstream WSEs alone are insufficient to explain the observed mismatch at peak flow and supports the interpretation* *that the adjustment of tributary inflow was the primary factor contributing to the improved consistency between the hydraulic model and SWOT observations for this event."*

2.5 Comment 5

Authors also failed to provide a critical discussion on interpretation of error metrics. In this manuscript RMSE values up to 0.54 m and biases as large as ±0.44 m is described as good agreement, yet these errors are significant relative to flood-stage variations and operational flood-monitoring requirements. Moreover, the manuscript also fails to justify whether these errors are driven primarily by SWOT measurement uncertainty or by hydraulic model limitations. Authors should explore where and why SWOT performs good or poor in a narrow channel, maybe they can explore spatial patterns in error related to channel geometry, curvature, or node averaging effects in a ~40 m wide channel.

Response:

We agree with the reviewer that the interpretation of error metrics such as RMSE and bias require careful contextualization. In the revised manuscript, we have strengthened the discussion to explicitly relate error magnitudes to the flood-stage variations. In particular, the water surface elevation varied by approximately 2 m during the flood event, and the reported RMSE values (up to 0.54 m) and biases (up to ±0.44 m) are now discussed relative to this event-scale variability and are not intended to represent absolute performance benchmarks for operational flood-monitoring applications.

In addition, the revised discussion places these results in the context of previous studies assessing the performance of satellite-based river water surface elevations and their use for hydraulic model evaluation or calibration. To further support this discussion, a new table (Table 3) has been added to the manuscript, summarizing reported RMSE values and their qualitative interpretation in previous satellite-based WSE studies across different river widths, missions, and hydraulic contexts. As summarized in Table 3, RMSE values on the order of 0.22-0.53 m are commonly reported and considered acceptable or reasonable in the literature, both for direct comparisons with in situ gauge measurements and for comparisons against hydraulic model simulations, particularly in large-scale or data-sparse basins (e.g., Michailovsky et al., 2012; Tourian et al., 2016; Shen et al., 2020; O'Loughlin et al., 2020; Zhou et al., 2023).

Importantly, the present study focuses on a narrow river (width ≈ 40 m), which is below the nominal detection threshold of conventional radar altimetry, and the evaluation is conducted during an extreme flood event. Under such conditions, uncertainties related to floodplain roughness, inundation extent, and channel-floodplain geometry are expected to be maximal, affecting both hydraulic model simulations and satellite observations. The fact that comparable error magnitudes are obtained in this particularly challenging hydraulic context further supports the consistency of the reported results with previous studies.

Following these revisions, RMSE and bias values are no longer used as stand-alone indicators of accuracy. Instead, they are discussed in relation to the observed flood-stage variability and compared with error ranges commonly reported in previous studies.

Lines 343-373 *"The results indicate a general consistency between SWOT observations and the WSEs simulated by the LISFLOOD-FP model across most of the analyzed cycles. The RMSE values, ranging from 0.22 m to 0.54 m, with biases between -0.30 m and 0.43 m were obtained, for river reaches with an average width of approximately 40 m. This width is significantly smaller than those investigated in most previous SWOT and radar altimetry validation studies (Table 3), which typically focused on rivers wider than 100 m and, in some cases, several kilometers wide. As reported by Domeneghetti et al. (2018), the performance of satellite-derived WSEs generally decreases as river width decreases, highlighting the challenging nature of the present case study.*

When interpreted relative to the hydrological signal of the event, the magnitude of the errors remains moderate. During the flood, WSE varied by approximately 2 m along the studied reach. In this context, the maximum RMSE of 0.54 m represents about 27% of the flood-stage amplitude, while the maximum bias of 0.42 m corresponds to approximately 22%. These values therefore remain limited when compared to the overall flood dynamics.

Beyond this event-specific perspective, the obtained WSE differences are consistent with those commonly reported in the literature for satellite-based river WSE observations and their application in hydraulic modelling. A selection of relevant studies summarized in Table 3 shows that satellite-derived river WSE errors, as quantified by RMSE, generally remain hydraulically informative for model evaluation and calibration across a range of observational accuracies.

For instance, Villadsen et al. (2016) synthesized past satellite missions' performance and proposed accuracy thresholds whereby errors below 0.30 m are considered "good" and values between 0.30 and 0.60 m considered "moderate". These criteria were subsequently adopted by Kittel et al. (2021b), who relied on the same benchmarks to assess the usefulness of satellite-derived WSEs for large-scale hydraulic modelling. Shen et al. (2020) later applied this framework in a river case study to interpret RMSE values and demonstrated that satellite-based WSE observations with good to moderate accuracy can effectively support hydraulic model evaluation and calibration. More recently, Zhou et al. (2023) showed that ICESat-2 and Sentinel-2-derived WSEs with RMSE values ranging from 0.25 m to 0.59 m enabled successful hydraulic model calibration, yielding validation RMSE values of approximately 0.36 m.

Importantly, the present results were obtained during an extreme flood event, a period during which uncertainties related to floodplain roughness (Pappenberger et al., 2005), inundation extent (Bates et al., 2014) (Aitken et al., 2024), and channel-floodplain interactions (Croke et al., 2013) are expected to be highest. These challenges are further exacerbated in narrow rivers, where satellite altimetry measurements are inherently more difficult to interpret.

Achieving WSE differences comparable to those reported in the literature under such conditions, and for a narrow river under the nominal detection limits of SWOT, further supports the hydraulic relevance of SWOT observations in this context. Overall, these findings indicate that, for the present case study, SWOT-derived WSEs provide hydraulically meaningful information for the evaluation of flood dynamics and hydraulic model performance, even in a narrow river and under high-flow conditions."

Table 3. Summary of relevant studies evaluating the performance of satellite-derived river WSE observations.

Reference	*Location*	*River width (m)*	*Mission*	*Reported RMSEs (m)*	*Comparison against*	*Qualitative assessment of reported RMSE*
Michailovsky et al. (2012)	*Zambezi River, Africa*	*< 80*	*Envisat*	*0.24–1.06*	*In situ gauges*	*"Good" < 0.40 m, "moderate" < 0.70 m, "bad" > 0.70 m*
Maillard et al. (2015)	*São Francisco River, Brazil*	*100–1000*	*Envisat, SARAL*	*0.02–1.63*	*In situ gauges*	*Acceptable precision*
Sulistioadi et al. (2015)	*Mahakam River, Indonesia*	*240–279*	*Envisat*	*0.69*	*In situ gauges*	*High accuracy*
Tourian et al. (2016)	*Po River, Italy*	*100–650*	*Envisat, Jason-2, SARAL/AltiKa, CryoSat-2, TOPEX/Poseidon*	*0.70–1.20*	*In situ gauges*	*Good correlation*
Shen et al. (2020)	*Han River, China*	*~400*	*Sentinel-3A, CryoSat-2*	*0.14–0.71*	*Hydraulic model*	*"Good" < 0.30 m, "moderate" < 0.60 m*
O'Loughlin et al. (2020)	*Congo River, Congo*	*~3900*	*ERS-2, Envisat*	*0.84*	*LISFLOOD-FP / In situ gauges*	*Relatively small RMSE and bias*
Jiang et al. (2021)	*Songhua River, China*	*~700*	*CryoSat-2, SARAL/AltiKa*	*0.40–0.50*	*Hydraulic model / In situ gauges*	*High precision*
Kittel et al. (2021)	*Zambezi River, Africa*	*Not specified*	*CryoSat-2, Sentinel-3*	*0.43–1.14*	*LISFLOOD-FP / In situ gauges*	*Satisfactory model performance; consistent with past studies*
Zhou et al. (2023)	*Yiluo River, China*	*10–500*	*ICESat-2, Sentinel-2*	*0.25–0.59*	*LISFLOOD-FP / In situ gauges*	*Reasonable accuracy*
Current study	*Du Gouffre River, Quebec*	*~ 40*	*SWOT*	*0.22–0.54*	*LISFLOOD-FP / In situ gauges*	*Accuracy consistent with satellite-based WSE performance reported for narrow rivers under flood conditions*

2.6 Comment 6

The main conclusion of this study that SWOT WSEs are "sufficiently accurate" for a ~40 m wide river, is already well reported in the recent past studies comparable or better *performance for rivers < 50–100 m. However, this manuscript failed to justify what new insight is added beyond confirming prior findings at one site. I suggest authors should critically investigate (1) whether measurements errors are systematic or site-specific, (2) whether performance is robust to geometry, slope, or hydraulic complexity, or (3) whether SWOT observations add information beyond what the hydraulic modelling framework (LISFLOOD-FP) already assumes.*

Response:

We agree with the reviewer that some recent studies have reported encouraging performance of SWOT-derived WSEs for rivers narrower than 50-100 m. However, studies specifically addressing narrow rivers focus on evaluating measurement performance or relative accuracy, often through comparisons with in situ data or other satellite products.

In contrast, the present study does not aim to provide an additional or generalized accuracy assessment of SWOT in narrow rivers. Rather, it adopts a hydraulic application perspective, examining how SWOT-derived WSEs can be used within an existing hydraulic modelling framework to support the analysis of a real extreme flood event in a narrow, poorly gauged river.

A key contribution of this work lies in demonstrating that SWOT observations provide independent, spatially distributed information that is not explicitly assumed by the hydraulic model. In the Du Gouffre River case, the temporal and spatial coherence between SWOT-derived WSEs and LISFLOOD-FP simulations across pre- and post-event cycles, together with independent validation at the downstream tide gauge, enabled the identification of an underestimation in upstream and tributary discharges during the flood peak. This illustrates how SWOT observations can complement hydraulic modelling by revealing inconsistencies in model forcing under data-scarce conditions, rather than merely confirming expected model behavior.

As such, the added value of this study is not limited to confirming that SWOT can perform reasonably well in a ~40 m wide river, but rather in demonstrating the practical usefulness of SWOT data for hydraulic studies, particularly for diagnosing flood dynamics and improving model consistency during extreme events. This distinction has been clarified throughout the revised manuscript, as detailed in the responses to the previous comments, which now explicitly frames the results as those of a site-specific case study focused on hydraulic application rather than a standalone performance evaluation.

We acknowledge that the extent to which these findings are systematic or transferable across different river geometries, slopes, and hydraulic complexities remains an open question. Addressing this will require future studies across multiple sites and events. Nevertheless, by documenting how SWOT observations add information beyond what is already prescribed in a hydraulic model during an extreme flood, this study complements existing validation efforts and helps advance the understanding of how SWOT data can be effectively used in hydraulic modelling of narrow rivers.

2.7 Comment 7

Increase the font size of Figures axis, and legends for ease of readability (e.g., Figure 4, Fig.5).

Response:

The font size of the axes labels and legends has been increased for improved readability.

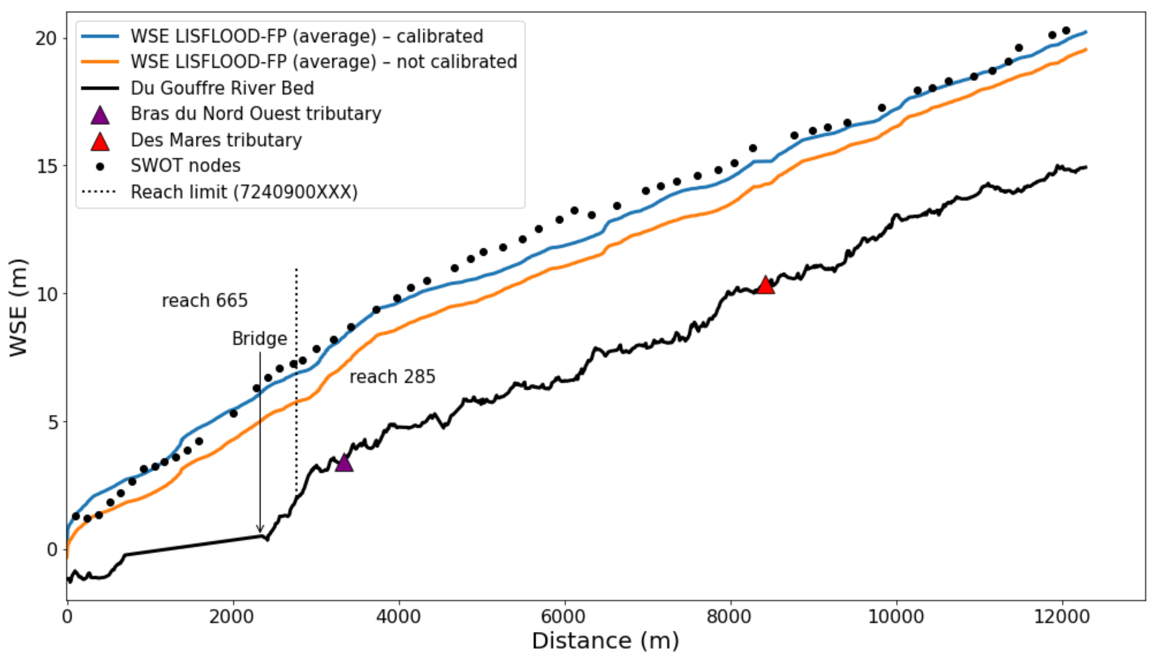

Figure 4: Longitudinal profile of WSE along the Du Gouffre River during the flood peak observed by SWOT (cycle 508, 1 May 2023). Black dots represent SWOT node-level WSE observations. The blue and orange lines show WSE simulated by the LISFLOOD-FP model using the calibrated and non-calibrated configurations, respectively. The black line corresponds to the riverbed elevation. The locations of the Bras du Nord-Ouest and Des Mares tributaries are indicated by colored symbols. Vertical dotted lines denote the limits between SWOT river reaches, and the location of a bridge structure is indicated.

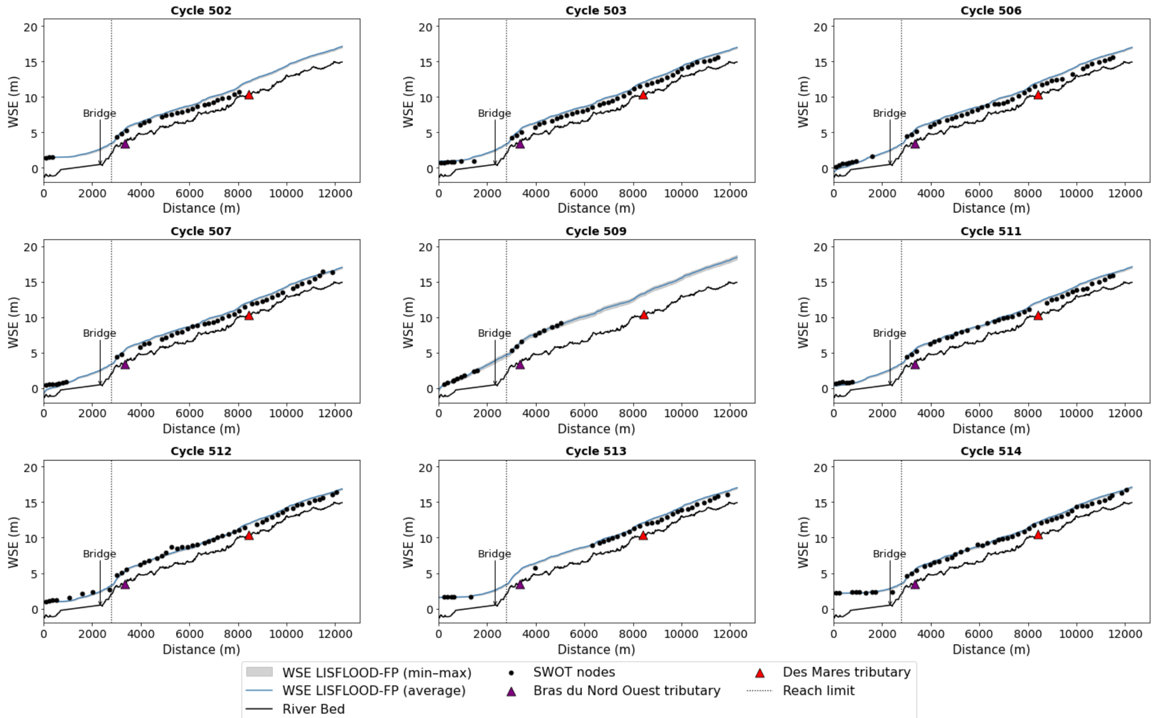

Figure 5: Longitudinal profiles of WSE along the Du Gouffre River for all SWOT overpasses before and after the 1 May 2023 flood peak, excluding cycle 508. For each cycle, SWOT node-level WSE observations (black dots) are compared with LISFLOOD-FP simulated WSEs, including the profile (blue line) and the envelope defined by the minimum and maximum discharges derived from rating-curve uncertainty (shaded area). The riverbed elevation is shown by the black line. The locations of the Bras du Nord-Ouest and Des Mares tributaries, the bridge structure, and the limits between SWOT reaches are indicated.

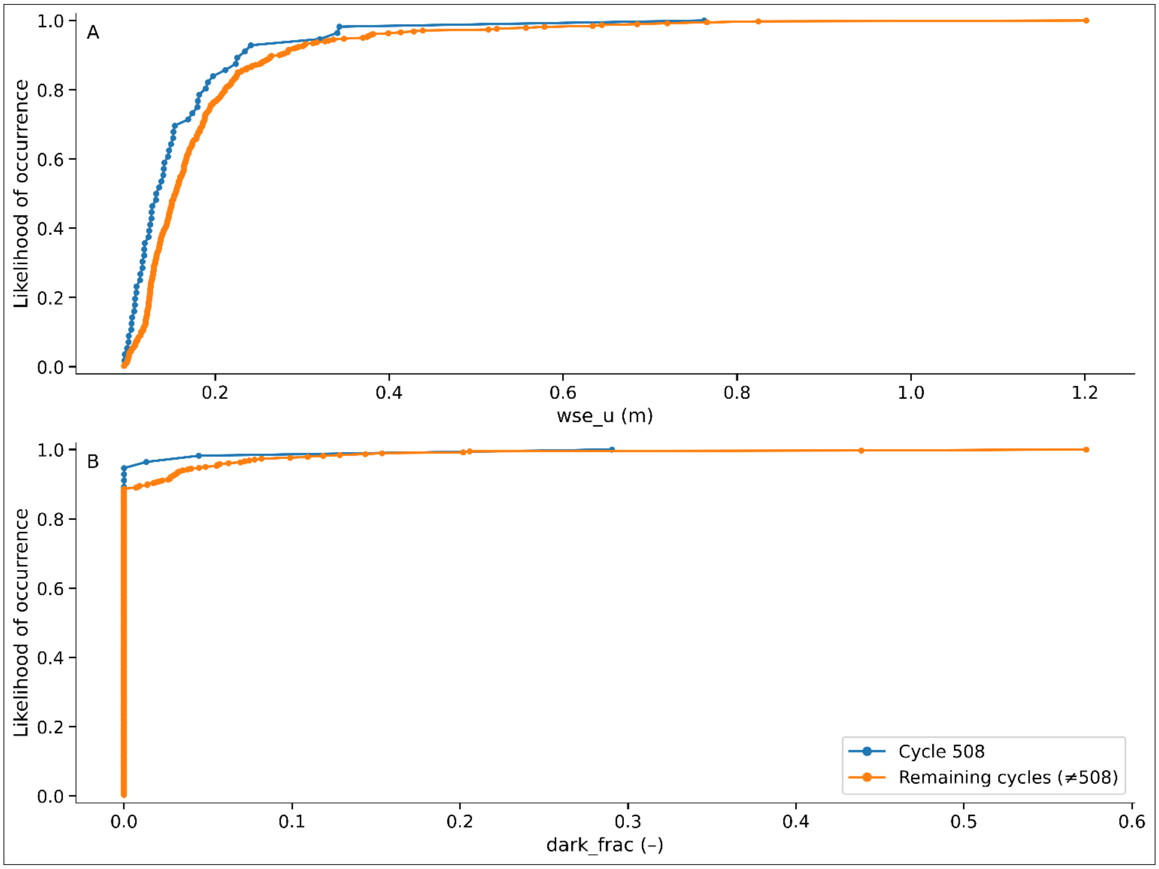

Figure 6: Likelihood of occurrence of SWOT node-level quality indicators for the Du Gouffre River: (A) wse_u (m) and (B) dark_frac (–). The flood-peak acquisition (cycle 508) is compared with all remaining SWOT cycles excluding cycle 508.

3 Response to Reviewer #3

We thank the reviewer for the positive assessment of the manuscript, for highlighting the relevance of the study, and for the constructive comments provided. This feedback has helped us clarify the scope of the study, refine the interpretation of the results, and improve the overall presentation of the manuscript. Our detailed responses to each comment are provided below.

3.1 Comment 1

The study covers only one river and one flood event, and most comparisons are with the hydraulic model rather than independent observations along the reach. I suggest the authors be more careful about broad claims on SWOT performance in rivers <50 m.

[Lines 27-28, 271-273] The abstract and discussion suggest the results apply broadly to narrow rivers, but the study only covers one ~40 m river and one flood. I suggest the authors clearly state that these findings are specific to this site and event.

Response:

This study is indeed based on a single river and a single flood event, and we agree with the reviewer that the results should not be generalized to all rivers narrower than 50 m. Please refer to our response to Reviewer 2, Comment 5, where this point is discussed in detail. In response, we have revised the abstract and the discussion to remove broad claims regarding SWOT performance in narrow rivers and to clearly state that the findings are specific to the studied ~40 m wide river and the analyzed flood event. The manuscript now consistently frames the analysis as a site- and event-specific case study, focusing on model-SWOT consistency under extreme flood conditions rather than on generalized performance assessment.

Please refer to our response to **Reviewer 2, Comment 1**, where the corresponding manuscript revisions are explicitly indicated.

3.2 Comment 2

[Lines 238–243, 290–295] For cycle 508, the authors adjusted the model discharge using SWOT data and then compared the model back to SWOT. I have two concerns: (a) please report bias and RMSE both before and after the adjustment, and (b) please clarify that this comparison shows model-SWOT consistency, not independent validation of SWOT accuracy.

Response:

We agree with the reviewer on both points. For cycle 508, the bias and error metrics values before and after the discharge adjustment have now been explicitly reported in the manuscript, allowing a transparent comparison of model-SWOT consistency prior to and following the adjustment. The manuscript has been revised to clearly state that this comparison is intended to illustrate model-SWOT consistency, rather than to provide an independent validation of SWOT measurement accuracy, since SWOT observations are used to inform the discharge adjustment during this cycle.

Please refer to our response to **Reviewer 2, Comment 2 (3)**, where the corresponding changes made to the manuscript are detailed.

3.3 Comment 3

[Lines 23–25, 271–273] The authors describe bias up to 0.44 m and RMSE up to 0.54 m as "good agreement" and "satisfactory accuracy," but do not justify why. Whether these errors are acceptable depends on the WSE amplitude during the flood. I suggest the authors discuss the error magnitude relative to the flood signal (for example, RMSE as a percentage of WSE range) to support these claims.

Response:

We agree with the reviewer that describing bias and RMSE values as “good agreement” or “satisfactory accuracy” requires explicit justification and appropriate contextualization. In response to this comment, the manuscript has been revised to relate the reported error magnitudes both to the flood-stage amplitude of the event and to performance ranges commonly reported in the literature for satellite-derived WSEs and hydraulic modelling.

Following these revisions, RMSE and bias values are no longer used as stand-alone indicators of accuracy. Instead, they are discussed in relation to the observed flood-stage variability and compared with error ranges commonly reported in previous studies. Consequently, expressions such as “good agreement” or “satisfactory accuracy” have been reformulated to better reflect this contextual interpretation.

Please refer to our response to **Reviewer 2, Comment 5**, this issue has been discussed in detail and the associated manuscript revisions are specified.

3.4 Comment 4

[Lines 195–198] Some nodes were removed manually, but the criteria are not clear. I suggest the authors (a) define objective rules for removal, (b) report how many nodes were removed per cycle, and (c) test how sensitive the results are to this step.

Response:

We agree with the reviewer that the criteria for node removal required clearer definition and justification. In response to this comment, the manuscript has been revised to (a) explicitly define objective and reproducible criteria for node removal, (b) report the number of nodes removed for each cycle, and (c) assess the sensitivity of the results to this filtering step.

(a) **Lines 233-237**

“In addition, manual node removal was applied only for three SWOT cycles (Table 2), following an objective consistency criterion. Specifically, nodes were excluded when the absolute difference between SWOT-derived WSE and the corresponding LISFLOOD-FP simulated WSE exceeded 1 m, while neighboring upstream and downstream nodes exhibited much smaller discrepancies and remained consistent with the expected longitudinal water surface profile. This procedure was applied to remove isolated outliers that were inconsistent with both the modeled WSE and the local spatial continuity of the river profile.”

(b) **Line 290**

Table 2. Comparison of WSEs observed by SWOT with the results of the LISFLOOD-FP model, for each SWOT cycle, together with a summary of manual SWOT node filtering. Initial nodes correspond to the number of nodes retained after automatic quality filtering, removed nodes indicate additional exclusions after the manual consistency filter, and the proportion removed represents the percentage of excluded nodes relative to the initial node count. Cycles 504-505 lack valid data and Cycle 510 was excluded after filtering.

SWOT cycle	*Bias (m)*	*RMSE (m)*	*Initial Nodes*	*Removed nodes*	*Proportion removed (%)*
502	*0.42*	*0.53*	*37*	*11*	*29.7*
503	*0.43*	*0.47*	*46*	*-*	*-*
506	*0.33*	*0.50*	*46*	*1*	*2.2*
507	*0.28*	*0.47*	*44*	*-*	*-*
508	*-0.30*	*0.54*	*56*	*-*	*-*
509	*-0.10*	*0.22*	*37*	*18*	*48.6*
510	*-*	*-*	*-*	*-*	*-*
511	*0.24*	*0.40*	*41*	*-*	*-*
512	*0.03*	*0.33*	*48*	*-*	*-*
513	*0.33*	*0.40*	*32*	*-*	*-*
514	*0.18*	*0.29*	*50*	*-*	*-*

(c) **Lines 420-433** *"Cycle 510 was excluded from the analysis because all available observations had an xovr_cal_q value of 2, indicating low-quality cross-over calibration. This shows the value of using the quality indicators supplied with the data to assess their reliability prior to any comparison with a model. Beyond this automatic filtering based on quality flags, a limited manual consistency filter was applied for three cycles, and the sensitivity of the results to this additional step was explicitly assessed. For cycles affected by isolated outliers, manual filtering led to a substantial reduction in the error metrics. This marked improvement reflects the strong influence of a small number of inconsistent nodes on aggregate error statistics, highlighting the disproportionate impact that isolated outliers can have on RMSE and bias estimates. For instance, for cycle 502, the RMSE decreased from 1.29 m to 0.53 m and the bias from 1.05 m to 0.42 m after manual filtering. Similarly, for cycle 509, the RMSE decreased from 6.51 m to 0.22 m and the bias from -5.5 m to -0.10 m, while the number of retained nodes was reduced accordingly. These cases illustrate that the large pre-filtering errors were primarily driven by a few extreme deviations rather than by a systematic mismatch between SWOT observations and model simulations. In contrast, for cycle 506, where only a single node was removed, RMSD and bias remained nearly unchanged (RMSD from 0.52 m to 0.50 m; bias from 0.35 m to 0.33 m). These results show that the manual filtering step mainly removes a few clearly inconsistent observations that strongly inflate the error metrics, rather than artificially*

improving the agreement between SWOT observations and the model. The overall conclusions of the SWOT-model comparison therefore remain robust. Future work should therefore focus on establishing explicit links between SWOT quality flags and the automatic identification of outliers. Such an approach would help reduced satellite-related uncertainties, improve reproducibility, and streamline the selection of reliable WSE observations."

3.5 Comment 5

[Figure 5] I find it hard to tell the two groups apart. I suggest using different line styles or separate panels.

Response:

Figure 5 was revised to improve readability by using distinct panels to clearly differentiate between the two groups.

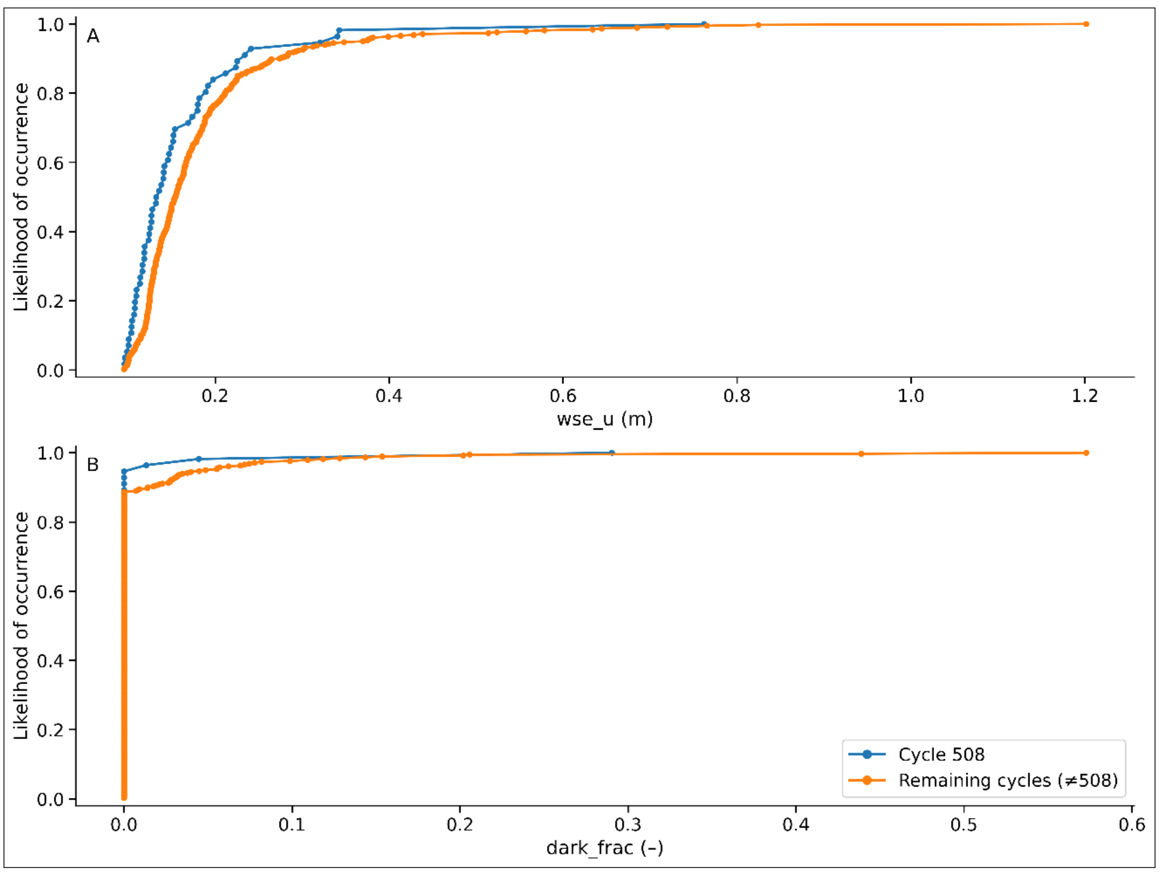

Figure 6: Likelihood of occurrence of SWOT node-level quality indicators for the Du Gouffre River: (A) wse_u (m) and (B) dark_frac (–). The flood-peak acquisition (cycle 508) is compared with all remaining SWOT cycles excluding cycle 508.