High-resolution mapping of pluvial flooding in ungauged agricultural catchments

Response to Reviewer 2

We sincerely thank Reviewer 2 for the time dedicated to evaluating our manuscript and for the thorough, constructive nature of their comments. Their rigorous analysis of our modelling framework, spatial scale constraints, and hydrological assumptions has provided a highly valuable perspective. The insights and additional literature suggested will significantly improve the scientific accuracy and clarity of this paper. Below, we address each point in detail. For clarity, the reviewer's comments are presented in plain text, and our corresponding responses are shown in blue.

Major comments

1. The analysis is based on a single catchment of only 111 ha, with an 84 ha nested sub-catchment. These are not two independent catchments with contrasting characteristics, but two highly dependent evaluation points within essentially the same hydrological system. This severely limits the possibility of assessing transferability across climatic, pedological, topographic, land-use, and catchment-scale conditions.

The very small catchment size may also explain why the two routing approaches provide almost identical results. The reported mean difference in concentration time is approximately seven minutes, which is close to one 5-minute computational time step. A larger or morphologically more complex catchment could produce very different conclusions regarding the relative performance of the routing methods.

The manuscript appropriately acknowledges the single-context limitation near the end, but this reservation is inconsistent with claims elsewhere that the approach constitutes a "transferable blueprint," is reliable for ungauged agricultural landscapes, or can support vulnerable landscapes globally. These statements should be substantially moderated throughout the Abstract, Results and Discussion, and Conclusions.

In this context, the authors could consider Soulis (2018), which showed that changing watershed conditions following wildfire altered not only runoff and peak discharge but also the apparent SCS-CN response, including a transition from complacent to standard behaviour. This illustrates that SCS-CN behaviour, and the relative performance of CN-based approaches can change markedly under different watershed conditions, even within the same catchment. The paper is particularly relevant to the statement that runoff partitioning is sensitive to local conditions and that results from one case cannot be generalized:

Soulis, K.X. (2018). Estimation of SCS Curve Number variation following forest fires. Hydrological Sciences Journal, 63(9), 1332–1346.
https://doi.org/10.1080/02626667.2018.1501482

We fully agree with the reviewer's critical assessment. The use of a single 111 ha catchment with an 84 ha nested sub-catchment does indeed represent two highly dependent evaluation points rather than contrasting systems, which intrinsically limits the immediate assessment of transferability across diverse climatic, pedological, and topographic conditions.

Focusing exclusively on the Chastre experimental catchment was a deliberate methodological choice driven by two main constraints. First, the methodological and theoretical explanations required for the distributed routing framework already resulted in a substantial manuscript length; introducing multiple contrasting catchments would have expanded the paper beyond manageable limits. Second, Chastre is our most comprehensively monitored and studied catchment. The availability of high-quality, high-resolution data at multiple reliable gauging points was essential to strictly evaluate the internal routing mechanics and hydrograph shapes presented in this study.

However, we recognize that some of our broader statements may have conveyed a level of generalizability that exceeds the evidential basis of a single-catchment study. The reviewer is correct that terms like "transferable blueprint" and claims of global operational readiness are overstated and unsupported by a single-catchment analysis. We will systematically revise the Abstract, Results and Discussion, and Conclusions to substantially moderate our claims.

Furthermore, we are grateful for the recommendation to incorporate Soulis (2018). We will include this reference in the Discussion section to further emphasize the sensitivity of the SCS-CN method to local watershed characteristics and evolving environmental conditions.

2. The performance of the runoff-generation component is far from satisfactory. Even the best configuration produces NSE = 0.17, RRMSE = 133%, and PBIAS of approximately 16%. Similarly, the median KGE values of 0.40 and 0.37 indicate only limited overall hydrograph reproduction, while errors during high-intensity and extreme events are frequently very large.

Against this background, the improvement from the lumped to the distributed Jain implementation is quantitatively small: NSE increases from 0.14 to 0.17, RRMSE decreases from 135% to 133%, and PBIAS changes from 19% to 16%. A statistically significant paired difference does not necessarily represent a hydrologically or operationally meaningful improvement. The authors should report an appropriate effect size and uncertainty intervals and discuss the practical magnitude of the improvement rather than relying mainly on the p-value.

Expressions such as “robust,” “reliable,” “acceptable runoff predictions,” “structurally robust,” and “critical operational value” are not supported by these performance levels. The text should distinguish more carefully between: 1) a technically functioning modelling workflow; 2) a statistically detectable difference between configurations; and 3) predictive performance sufficient for practical flood-risk decisions.

The current results primarily support the first of these.

We agree with the reviewer’s rigorous evaluation of our performance metrics. As discussed in the manuscript, the modest overall performance is fundamentally tied to the standard runoff production module. We deliberately selected this simplified, standard production method to maintain parsimony and test a framework that could be easily implemented in ungauged contexts. However, we acknowledge that this choice bottlenecks the overall model performance. We had already tested the upper performance boundary of the routing module by conducting a theoretical experiment where the simulated runoff volume was artificially forced to match the observed volume (i.e., isolating the routing module by assuming perfect runoff production). Under these conditions, the median KGE for the distributed routing approach rises to 0.685. It is important to note that our results approach this theoretical ceiling. As highlighted by the reviewer, the overall performance metrics are pulled down by specific events characterized by high rainfall intensities over short durations (e.g., Events 7, 23, and 35, Table B1). This illustrates that while the routing module itself performs adequately and generally captures the timing and shape of the hydrograph well, it struggles with short, intense rainfall, and its true potential is masked by the limitations of the standard, static SCS-CN production module.

We agree with the reviewer's point regarding the small quantitative improvement between the lumped and distributed implementations. We will adjust our analysis to discuss the practical magnitude of this difference.

We appreciate the reviewer's clear distinction between a "technically functioning modelling workflow" and "predictive performance sufficient for practical flood-risk decisions". Accordingly, we will systematically revise the terminology throughout the manuscript to ensure our claims are appropriately cautious and strictly aligned with our results. We agree that using terms such as "robust" or "reliable" may overstate the absolute predictive accuracy of our "uncalibrated" outputs, and we will systematically temper this terminology throughout the manuscript to ensure our claims remain appropriately cautious.

Nevertheless, it is important to note that high predictive performance is entirely achievable in this catchment when a more complex, locally calibrated production module is employed. For instance, in a recent study on the exact same 84 ha sub-catchment, Baert et al. (2026; under review) utilized a threshold-controlled approach where event-specific initial abstraction ratios were optimized against observed rainfall-runoff responses and subsequently generalized via a regime-based clustering strategy. Their approach yielded an impressive NSE of 0.85 over 46 events. This demonstrates that while the catchment's hydrological behaviour could be further optimized through more thorough local calibration, our primary objective was to rigorously evaluate the baseline performance of a simple and low-parameterized approach specifically designed for data-scarce environments.

3. The improvement obtained by calculating runoff separately for spatial units rather than using an area-weighted composite CN is an important result. However, it is already well established that, because of the nonlinear form of the SCS-CN equation, runoff calculated from a composite CN can differ substantially from the area-weighted runoff calculated for heterogeneous subareas.

The Introduction should explain this issue explicitly and engage with literature that investigates the effect of watershed heterogeneity on CN estimation and runoff prediction. Two particularly relevant studies are:

Soulis, K.X. and Valiantzas, J.D. (2012). SCS-CN parameter determination using rainfall-runoff data in heterogeneous watersheds – the two-CN system approach. Hydrology and Earth System Sciences, 16, 1001–1015. https://doi.org/10.5194/hess-16-1001-2012

Palli-Gravani, S., Soulis, K.X. and Wałęga, A. (2026). Simplified approach to considering watershed heterogeneity in direct runoff estimation using SCS-CN model. Hydrological Sciences Journal. https://doi.org/10.1080/02626667.2026.2628736

The first study explains theoretically and empirically how spatial variability affects the CN–rainfall relationship and why a single composite CN can generate important errors. The second provides a recent multi-catchment and synthetic-data assessment of how explicitly representing the spatial distribution of runoff-producing areas improves CN estimation. These studies would strengthen both the Introduction and the interpretation of the distributed-versus-lumped comparison. The present improvement should be presented in relation to this existing understanding rather than as an isolated finding.

We thank the reviewer for highlighting the importance of this result, which is indeed one of the key points we aim to emphasize in this study. We agree that the divergence in runoff volumes between a distributed calculation and a composite CN is a well-established mathematical consequence of the non-linear SCS-CN equation. We did not intend to present this divergence as a novel discovery, but rather as a necessary structural validation of our distributed framework. However, we acknowledge that the current framing may inadvertently suggest otherwise, and we will carefully revise the manuscript's narrative to correct this impression.

To address this, we will restructure the Introduction and Discussion to integrate the suggested references (Soulis and Valiantzas, 2012; Palli-Gravani et al., 2026). This will ensure the improvement observed in our distributed implementation is explicitly positioned as a confirmation of the existing scientific consensus on watershed heterogeneity, rather than a novel finding.

4. I am not convinced that all components of the modelling framework are physically meaningful at a 1 m resolution. The DEM has metre-scale information, but several other inputs, including hydrologic soil groups, CN values, crop classes, Manning coefficients, and channel parameters, are derived from much broader classifications or coarser datasets. Running the model on a 1 m grid does not necessarily imply that runoff generation and hydraulic properties are known at that scale.

The manuscript needs a systematic resolution-sensitivity analysis. Results should be compared at several resolutions, for example 1, 2, 5, and 10 m, considering for instance: delineated flow paths; concentration and travel times; peak discharge; hydrograph performance; computational time and memory requirements; and sensitivity to the imposed minimum slope.

At present, the argument for selecting 1 m is based largely on literature and qualitative reasoning. Appendix A appears to examine the effect of resolution on local velocity classifications, but it does not demonstrate that 1 m provides superior observed hydrograph or peak-flow predictions in this catchment.

The authors should also quantify the error introduced by rounding CN and travel-time values to integers for the grouped convolution procedure.

We thank the reviewer for raising this fundamental modelling question and for suggesting this rigorous evaluation. As noted, although the computational grid is defined at 1 m resolution, not all input datasets are available at this scale. Land-use and soil information are derived from datasets with spatial resolutions of 2 m and 10 m, respectively, with land use being used to derive most hydrological parameters (e.g., CN and Manning coefficients). We therefore do not consider this approach to create an artificial illusion of precision. Moreover, the 2 m land-use dataset provides a much more detailed representation of landscape heterogeneity than coarser products, allowing the identification of narrow linear features such as hedgerows, grass strips, wooded buffers, roads, and other obstacles that can significantly influence runoff pathways and hydrological connectivity.

The primary purpose of the 1 m grid is to accurately represent topographic controls on flow routing, connectivity, and concentration pathways. The fine spatial discretization should therefore be viewed as a computational framework for capturing flow pathways and topographic features identified in the high-resolution DEM, rather than as an indication that all hydrological properties are known at metre-scale resolution. While the effective resolution of some parameters remains constrained by the underlying thematic datasets, the 1 m grid enables these parameters to interact with a highly detailed representation of topography.

This specific choice is heavily supported by previous research conducted in the same 111 ha Chastre catchment. Cantreul et al. (2018), cited in the manuscript, systematically evaluated the effect of DEM resolution (ranging from 0.25 m to 10 m) on sediment connectivity. They demonstrated that at resolutions of 5 m or coarser, there is a severe loss of information regarding straight flow paths, and linear landscape features are no longer visualized. Conversely, they identified 1 m as the optimum DEM resolution for this specific environment. A 1 m grid accurately captures essential connectivity features, field limits, and disconnected areas (e.g., zones of water stagnation) without the prohibitive computational costs of sub-half-meter grids. As documented in the literature cited in the manuscript (e.g., Veeck et al., 2020; Alexopoulos et al., 2024), finer spatial resolutions are essential for capturing realistic flow connectivity and maximizing overall model accuracy.

To quantitatively support our approach and address the reviewer's request, we conducted a systematic resolution-sensitivity analysis comparing 1 m, 2 m, 5 m, and 10 m grids across the 39 observed events (Table 1), which will be included in a new Supplementary Material section. The analysis reveals that coarsening the DEM artificially accelerates flow routing due to the smoothing of the topography. This artificial acceleration leads to severe overestimations of peak discharge (Qp) at 5 m (+25%) and 10 m (+28%). While the 2 m resolution cuts computational runtimes and slightly improves specific routing timing metrics (reducing the median relative error on Tc to 16%), it begins to introduce an overestimation of peak flows (+6%) and degrades the overall hydrograph shape performance (median KGE of 0.30, compared to 0.35 at 1 m). Because capturing the most accurate hydrograph shape and preserving the physical representation of topographic flow pathways are paramount to this research, and in accordance with the literature, we retain the 1 m resolution for this framework, especially since its computational cost remains highly manageable (4.2 min). We will, however, note in the discussion that a 2 m grid could serve as a viable operational compromise, particularly for wider regional applications. Finally, we will take this opportunity to correct a methodological error in the current manuscript regarding the KGE calculation, where negative values were previously capped at -1 for display purposes, thereby artificially lowering the reported performance across all resolutions.

Table 1 – Comparison of hydrological performance metrics across spatial resolutions for the SCS transfer method. Metrics include the time of concentration (Tc), time to peak (Tp), peak discharge (Qp), runoff volume (V), the KGE between observed and simulated hydrographs, and the total computation time required to simulate the 39 rainfall–runoff events.

Metrics – Median (Stdv)	1 m	2 m	5 m	10 m
Relative error on Tc (%)	38 (77)	16 (64)	-2 (55)	-7 (52)
Relative error on Tp (%)	16 (59)	10 (48)	3 (46)	1 (46)
Relative error on Qp (%)	-7 (121)	6 (132)	25 (143)	28 (146)
Relative error on V (%)	5 (194)	2 (189)	-2 (184)	1 (188)
KGE	0.35 (1.5)	0.30 (1.6)	0.23 (1.5)	0.21 (1.6)
Computational time (min)*	4.2	2.3	1.3	0.7

* Simulations carried out on a computer with 32 GB RAM and a CPU of 4.

The sensitivity analysis also highlights the interaction between spatial resolution and the imposed minimum slope threshold. As requested, we evaluated the sensitivity of Tc to the minimum slope threshold across all resolutions. Across both catchments, the acceptable range for the minimum slope threshold spans from 0.3% to 5%. This specific interval yields simulated Tc values that align closely with both direct field observations and the empirical lag time relationship proposed by Mockus (1972) (Figs. 1 and 2). In the 84 ha catchment, observed Tc ranges from 15 to 85 minutes (median: 50 minutes). At 1 m resolution, reproducing this median would require imposing a minimum slope of approximately 0.02 m/m (Fig. 2), which is inconsistent with field observations and topographic analyses indicating thalweg slopes of around 0.005 m/m. In line with our partially calibrated modelling philosophy, we retained a physically realistic minimum slope threshold rather than artificially inflating it to optimize timing metrics. We acknowledge that this choice contributes to the slight overestimation of temporal metrics at 1 m resolution, but it ensures the model remains grounded in the actual catchment topography. We will clarify this reasoning more explicitly in the Methods section of the revised manuscript.

The spatial analysis of surface runoff concentration pathways is addressed in point 6.

Finally, regarding the grouped convolution procedure, we quantified the error introduced by rounding the CN and travel time values to integers. Tests conducted on a subset of 29 events demonstrate that the rounding process introduces negligible error while providing immense computational benefits. The mean absolute difference in predicted peak discharge between the rounded and unrounded methods is only 0.50 ± 0.81 l s^{-1}, and the mean NSE between the two resulting hydrographs is 0.999. In contrast, processing these 29 events without rounding required 138 minutes of computational time, whereas the rounded grouping procedure executed in a few minutes. We will shortly incorporate these validation metrics into the methodology section to robustly justify this computational optimization.

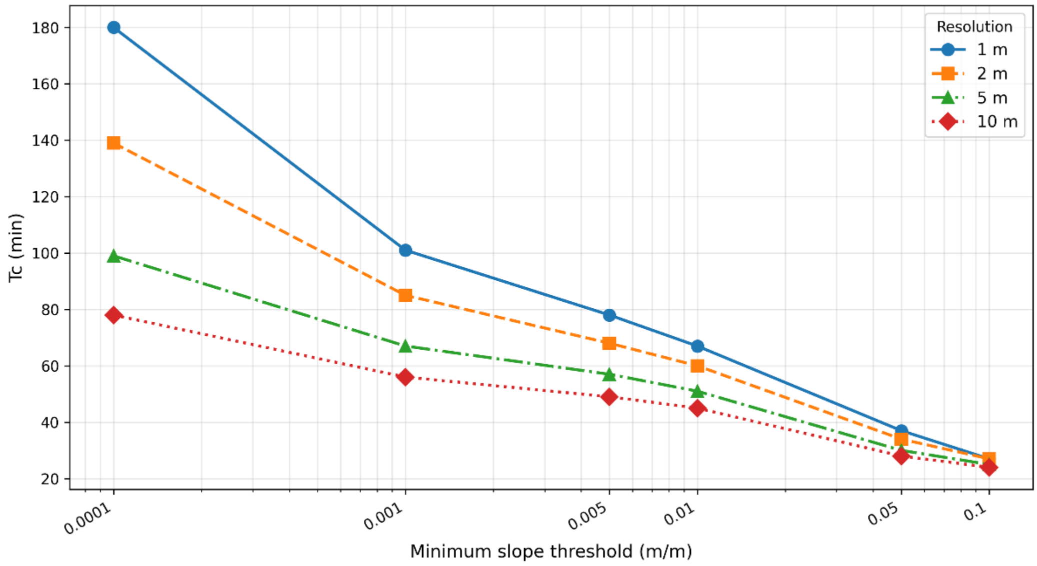

Figure 1 – Sensitivity of time of concentration to minimum slope threshold and resolution in the 111 ha catchment.

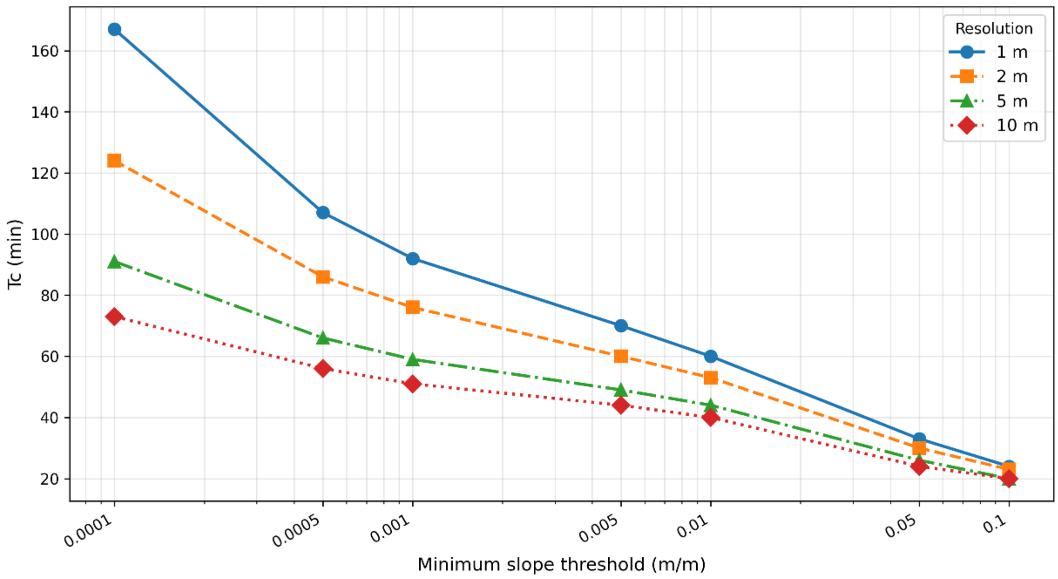

Figure 2 – Sensitivity of time of concentration to minimum slope threshold and resolution in the 84 ha catchment.

5. The claim that the approach can be applied without calibration requires reconsideration. Several parts of the routing framework appear to involve catchment-specific adjustment:

- the stepwise relationship between contributing area and flow depth is explicitly described as site-specific and calibrated using field observations;

- the minimum slope of 0.005 m m^{-1} was selected because it produced concentration times consistent with observed values and another empirical equation;

- the adjustment of travel times for extreme events was evaluated using preliminary sensitivity tests in this catchment, but the results are not shown;

- the transition thresholds between hydraulic regimes are empirical.

These choices may be reasonable for the experimental catchment, but they contradict repeated statements that the framework is uncalibrated and directly transferable to ungauged catchments. The authors should clearly separate:

1. parameters taken unchanged from published sources;

2. parameters inferred from local field observations;

3. parameters selected or tuned against the observed hydrographs; and

4. parameters used without any evaluation.

Ideally, calibration and validation events should be separated, or a leave-one-event-out or split-sample analysis should be undertaken. Otherwise, the method should be described as locally parameterized or partially calibrated rather than uncalibrated.

We thank the reviewer for this valuable remark. We concede that "partially calibrated" or "locally parameterized" is a more transparent and accurate description of our methodology.

To address this, we will systematically replace the term "uncalibrated" throughout the manuscript to correctly reflect the partially calibrated nature of our framework. Furthermore, to ensure full transparency and reproducibility, we will explicitly categorize all model parameters within the Methodology section according to the reviewer's suggestion. Specifically, we will detail which parameters were taken unchanged from published sources, inferred from local field observations, tuned against observed hydrographs, and used without formal evaluation.

Regarding the adjustment of travel times for extreme events, the results are not presented here, as these analyses rely on a limited sample of only four extreme events, which is insufficient to draw sound conclusions.

6. The concentrated flow paths shown over the background imagery require validation. This is especially important in the northern cultivated part of the catchment, where crop rows, tractor tracks, wheel ruts, tillage direction, field boundaries, temporary ditches, roads, and agricultural operations may redirect runoff independently of the bare-earth DEM.

The mapped paths appear relatively natural and topographically controlled, but the agricultural fields are heavily managed and their surface configuration changes over time. Were field surveys undertaken to confirm the location of concentrated flow, ephemeral gullies, muddy-flow traces, or erosion features? Were any mapped pathways compared with UAV imagery, event photographs, observed erosion patterns, or farmer reports?

Without independent spatial validation, good or moderate agreement at one or two outlets does not demonstrate that the internal drainage pathways or mapped hotspots are correct. This is a central issue because the main claimed added value of the study is the spatial identification of critical locations.

We thank the reviewer for this critical point. We agree that spatial validation of internal drainage pathways is essential to support our claims regarding the identification of critical locations, as good agreement at the outlet does not guarantee correct internal routing. The spatial distribution of concentrated flow paths in this specific 111 ha catchment has been documented and partially validated in previous studies (Cantreul et al., 2018, 2020; Herpoel et al., 2025) using field surveys and aerial photography. These observations confirm that while minor flow paths can shift inter-annually due to agricultural practices such as tillage, the main concentrated flow paths remain topographically fixed and are often marked by ephemeral gullies following intense storms (Fig. 3).

Figure 3 – Aerial view of the Chastre headwater catchment following an intense storm on May 29, 2026. © Gilles Swerts.

To address the reviewer's specific concern regarding the northern cultivated area, we mapped the spatial dispersion of flow paths to create an uncertainty layer (Fig. 4). This analysis reveals a high variability of minor pathways in this northern section, highlighting the mobile nature of these secondary routes and confirming the boundary effects pointed out by the reviewer. Furthermore, we assessed the accuracy of the mapped pathways against observed gully networks following a recent intense storm on May 29, 2026 (Fig. 5). This comparison shows that the 1 m and 2 m resolutions provide a more precise alignment with the observed ephemeral gullies compared to the 5 m and 10 m resolutions, which also provides empirical support for our selected grid size.

Given the current length constraints of the manuscript, we will compile these spatial validation analyses into a new Supplementary Material document. This will include the flow path uncertainty maps, the resolution-sensitivity comparison regarding gully location, and the recent field observations. We will add a focused paragraph in the Results and Discussion section of the main text referencing this supplementary validation to substantiate our spatial predictions without overburdening the article.

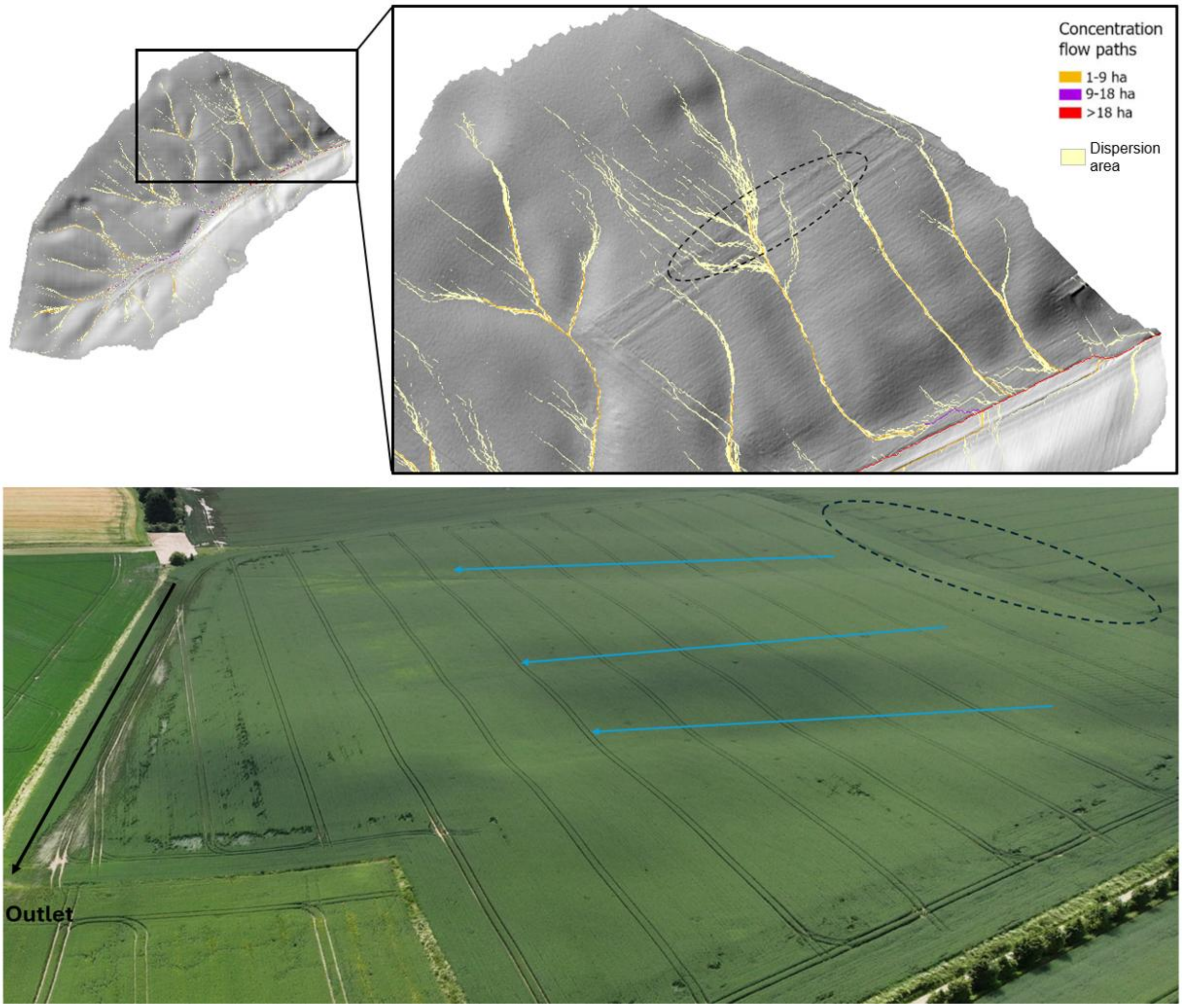

Figure 4 – Top: Spatial dispersion and mapped location of simulated flow paths. Bottom: Aerial view of the northern hillslope of the Chastre catchment following the May 29, 2026 storm. The image highlights the position of the three main concentrated flow axes (blue) converging into the primary thalweg (black). Flow paths within the dotted area exhibit a diffuse and mobile pattern. © Gilles Swerts.

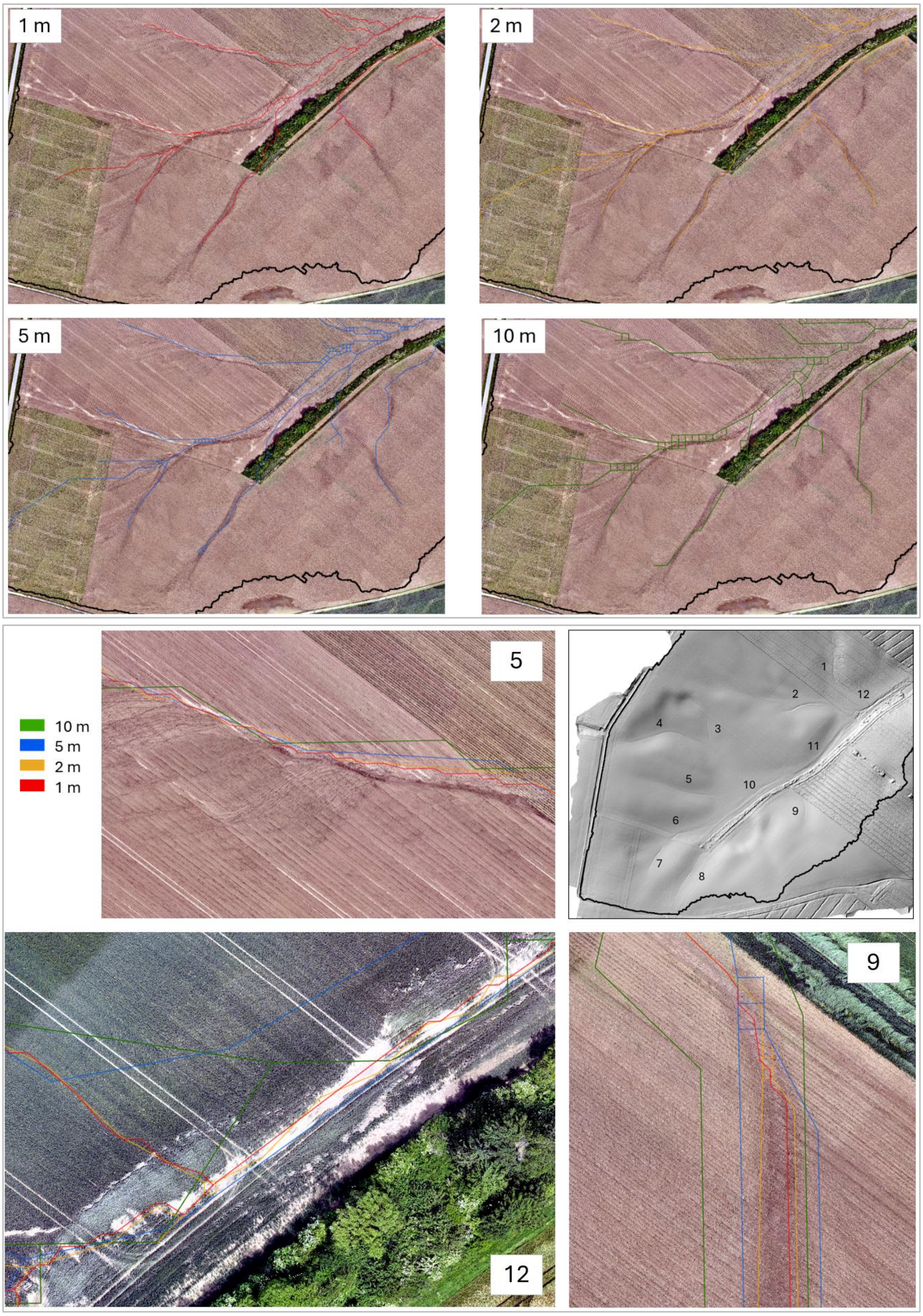

Figure 5 – Simulated concentrated flow networks extracted at four different spatial resolutions (1 m, 2 m, 5 m, and 10 m) based on the 2021-2022 DEM. Location of 12 ephemeral gullies observed following the May 29, 2026 storm, mapped on the post-event DEM (captured May 30, 2026).

7. The manuscript repeatedly refers to "flood hazard maps," although the main output is peak discharge along predefined concentrated flow paths. Flood hazard normally involves some combination of inundation extent, water depth, velocity, duration, probability, and potentially consequences for exposed receptors.

Peak discharge is certainly a useful hydrological indicator, but it is not by itself a complete representation of flood hazard. The authors should either: reframe the output as spatial peak-discharge or concentrated-flow screening; or couple it with a hydraulic or terrain-based procedure that estimates water depth, velocity, or inundated area.

The practical relevance of flood hazard in this specific 111 ha agricultural basin should also be better demonstrated.

We thank the reviewer for this valid conceptual clarification. We concede that mapping peak discharge alone represents a hydrological indicator rather than a complete hydraulic hazard. However, it is important to contextualize this approach within the regional framework of our study. Under Walloon legislation, pluvial flood hazard is officially evaluated based on the combination of peak discharge and the rainfall event's return period. Furthermore, while the maps display peak discharge for visual simplicity, the proposed modelling framework computes the full hydrograph for every target pixel along the concentrated paths. This provides flow duration and temporal dynamics, allowing end-users to derive hazard indices according to their own specific operational definitions.

Regarding the practical relevance of evaluating flood hazards in a seemingly small 111 ha agricultural basin, we must emphasize that in Wallonia, and particularly within the Belgian loess belt, rapid pluvial flooding and associated muddy flows constitute the predominant flood risk, significantly surpassing fluvial (river) flooding. Driven by the industrialization of agricultural practices and soil compaction, this specific surface runoff hazard affects approximately one-third of exposed residential properties in the region. The Chastre experimental catchment is representative of these vulnerable loess landscapes, which continuously face severe erosion and runoff issues, as recently demonstrated again by the intense local storm of May 29, 2026 (Fig. 3). To address the reviewer's concern, we will explicitly clarify the semantics of our screening maps and expand the Study Area section to better demonstrate the severe socioeconomic relevance of pluvial flooding originating from these small headwater catchments.

8. A central justification for the work is its potential application in ungauged regions. However, no evidence is provided that a 1 m iterative implementation is computationally feasible beyond this exceptionally small catchment.

The manuscript should report: the hardware used; preprocessing and simulation times; memory requirements; the number of active pixels and temporary outlets; the reduction achieved by grouping identical CN–travel-time combinations; and how runtime scales with catchment area and drainage-network length.

An application or benchmark for a larger area would considerably strengthen the study. Without such evidence, statements about wider operational use remain speculative.

We thank the reviewer for this crucial comment. We fully agree that demonstrating computational feasibility is essential to support our claims regarding the framework's operational potential for ungauged regions. We will add a brief dedicated subsection in the methodology detailing the computational requirements. The simulations were performed in Python 3.11 on a standard workstation equipped with a 4-core CPU and 32 GB of RAM. For the 111 ha Chastre catchment at a 1 m resolution, simulating the 39 events at the main outlet required only 1.4 minutes of preprocessing and 2.8 minutes of processing (Table 1). For the application along the concentrated flow paths, the catchment comprises 1,110,000 active cells draining towards 8,689 temporary target outlets. Simulating a single storm for all these target pixels takes only 2 minutes. This high efficiency is achieved notably through our grouping procedure, which drastically compresses the grid data by reducing approximately 1,065,000 unique pairs (CN, travel time) into roughly 550 pairs for the mathematical convolution.

To address the reviewer's request for a larger benchmark, we note that the framework had already been successfully applied to four larger Walloon catchments (Table 2) from the AQUALIM monitoring network (Flobecq, Upigny, Modave, and Forêt; Fig. 6) prior to this review. These results were not included in the current manuscript to keep the paper focused, as these larger basins exhibit continuous streamflow requiring baseflow separation procedures that would lengthen the article. However, we will include a summary of these computational benchmarks in the revised supplementary material to prove scalability.

Table 2 – Characteristics of the catchments within the AQUALIM network and the associated calculation times for applying Hidropixel to the outlet. LHP = Longest Hydraulic Path; RHW length = total length of the Walloon hydrographic network within the catchment.

Characteristics	Flobecq (L7590)	Upigny (L6260)	Modave (L7860)	Forêt (L7600)
Surface (ha)	1,500	1,800	3,000	4,000
LHP (km)	7.9	10.4	16.8	17.4
RHW length (km)	23	26	25	52
Active cells (.10^6)	15	18	30	40
Number of pairs (CN, Tv)	5,800	6,600	9,000	6,600
Simulation times (min)	0.9	1.2	2.4	2.5

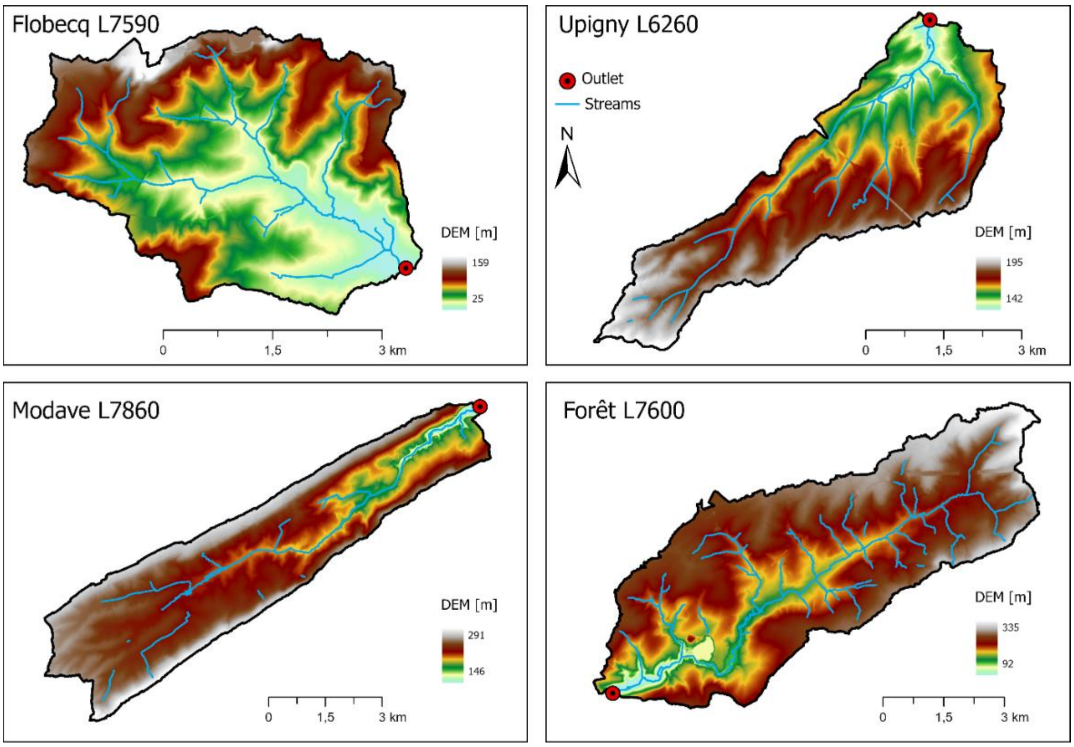

Figure 6 - Description of the AQUALIM catchment areas, including watercourses, topography and outlets.

As an example, applying the model to the much larger Flobecq catchment (using a 0.5 ha drainage threshold) generates 115,650 temporary outlets and contains 15 times more active cells than Chastre. Simulating one storm across this vast 1 m grid required 198 minutes.

Ultimately, these larger-scale applications demonstrate the framework's computational viability. Runtime is primarily driven by three interacting factors: the total active catchment area, the chosen drainage-network extraction threshold (which dictates the number of temporary outlets), and the total number of distinct (CN, travel time) combinations. As shown in Table 2, the quantity of these unique combinations inherently grows with both the overall surface area and the broader range of travel times found in larger basins. Consequently, when applying the framework to wider regional scales, both the drainage threshold and the spatial resolution must be carefully considered. This strategic parameterization is essential to maintain the hydrological relevance of the simulations while preventing computational bottlenecks caused by an excessive number of grid cells.

9. The authors correctly state that the 25- and 100-year simulations are exploratory scenarios rather than validated predictions. This qualification should be maintained consistently throughout the manuscript. Statements that the model "demonstrates" the expansion of the active flow network or identifies nonlinear synchronization under extreme events are stronger than warranted, because these patterns are generated by the assumed model structure and travel-time adjustment rather than independently observed. The reported increase of more than 90% along the main thalweg should be described as a modelled scenario outcome, not as an empirically established hydrodynamic response.

We agree with the reviewer's observation. To rectify this, we will review and revise the terminology throughout the manuscript.

10. The manuscript states that the Jain initial abstraction formulation depends on rainfall intensity. However, the presented equation depends on total rainfall depth, P, and potential retention, S, not on rainfall intensity or the temporal distribution of rainfall. This distinction is important because the manuscript subsequently identifies rainfall intensity as the principal unresolved driver of model error.

The terminology should be corrected throughout. The Jain formulation may make initial abstraction dependent on storm depth, but it does not resolve within-event rainfall intensity. Similarly, the statement that the SCS-CN framework was "empirically calibrated for 24-hour events" should be carefully checked and qualified. The fundamental problem here is that the method is an event-total runoff equation without an explicit infiltration-rate or within-storm intensity representation.

We thank the reviewer for this precise observation. We fully agree that the Jain formulation depends on total rainfall depth and potential retention, rather than rainfall intensity. We will systematically correct the terminology throughout the manuscript. Accordingly, we will explicitly attribute the model's errors during short storms to the structural inability of the framework to resolve within-event temporal intensity. We will also adjust the statement that the SCS-CN method is "empirically calibrated for 24-hour events," defining it instead strictly as an event-total volume method.

11. The discussion of antecedent conditions should be strengthened and better supported by the relevant literature. The finding that model errors are not significantly correlated with 5-day antecedent rainfall is consistent with a growing body of research showing that antecedent rainfall is often a poor surrogate for the actual watershed conditions controlling runoff generation and that the traditional use of antecedent moisture classes may not necessarily improve SCS-CN applications.

The authors should therefore place their results in the context of previous studies that have questioned the effectiveness of antecedent rainfall indices as predictors of runoff response. In particular, the papers by Soulis and Valiantzas (2012), Soulis (2018), and Palli-Gravani et al. (2026) provide useful evidence that watershed heterogeneity and changing watershed conditions can exert a stronger influence on runoff estimation than simple antecedent-rainfall classifications. These studies also illustrate that runoff response cannot always be adequately represented through conventional antecedent-moisture adjustments.

Rather than interpreting the lack of correlation with P5 as an isolated result, the authors should discuss whether their findings support the broader view that antecedent rainfall is not necessarily an appropriate indicator for adjusting CN values in event-based applications. Such a discussion would place the results within the existing literature and provide a stronger justification for the modelling choices adopted in the manuscript.

We agree with the reviewer's assessment regarding the limitations of the P5 index. As noted in the manuscript, updated guidelines explicitly discourage its use due to a lack of statistical correlation with observed CN variations (Hawkins et al., 2008; USDA-NRCS, 2010). We will expand the discussion section to integrate the additional references suggested by the reviewer, thereby providing an even more robust theoretical justification for avoiding P5-based adjustments.

12. There is considerable repetition between the Introduction, methodological overview, and detailed method description. The manuscript could be shortened and made more focused by reducing repeated explanations of Hidropixel, the two modelling phases, and the limitations of static CN methods.

More importantly, the conclusions should remain closely aligned with the evidence. The study demonstrates a technically interesting implementation and identifies important limitations of the production model. It does not yet demonstrate a generally reliable or transferable flood-hazard tool. Statements such as "valuable tool for risk-informed land management," "successfully captured," "structurally robust," and "critical to support proactive land management globally" should be revised to reflect the proof-of-concept nature of the analysis.

We thank the reviewer for this relevant comment. Removing this redundancy will effectively help streamline the manuscript. We will delete the repetitive sections and systematically temper our assertions to ensure all claims remain suitably cautious and strictly supported by our results.

<u>Minor comments</u>

- The event-selection sentence stating that events were retained "only if they met this rainfall depth" appears incomplete. The rainfall threshold should be specified.

We thank the reviewer for catching this omission. We will update the manuscript to explicitly specify the exact rainfall depth threshold used for the event-selection process to ensure the sentence is complete.

- Please state clearly how many events were available at each monitoring point. Events from the nested outlets should not be treated as independent observations in statistical analyses.

The exact number of events available at each monitoring point is detailed in Table B of Appendix B. Furthermore, we acknowledge the reviewer's valid point regarding the nested nature of the outlets. We will clarify in the methodology how we account for this lack of statistical independence when interpreting the performance across the different stations.

- The smoothing procedure and numerical implementation used to identify the hydrograph inflection point should be fully described. The uncertainty of the observed concentration time should also be quantified where possible.

We thank the reviewer for this comment. We will update the text to specify the exact smoothing algorithm and window size applied to the raw discharge time series prior to computing the derivatives. Regarding the uncertainty of the observed concentration time, we will clarify that our algorithmic approach ($d^2Q/dt^2 = 0$) was specifically implemented to ensure an objective, systematic identification of the inflection point, thereby eliminating the subjectivity and human error inherent to manual graphical extraction (Fig. 7). While providing a formal quantitative uncertainty bound is challenging for this deterministic mathematical extraction, we will add a brief sentence in the methodology acknowledging that the resulting Tc remains inherently sensitive to the selected smoothing parameters.

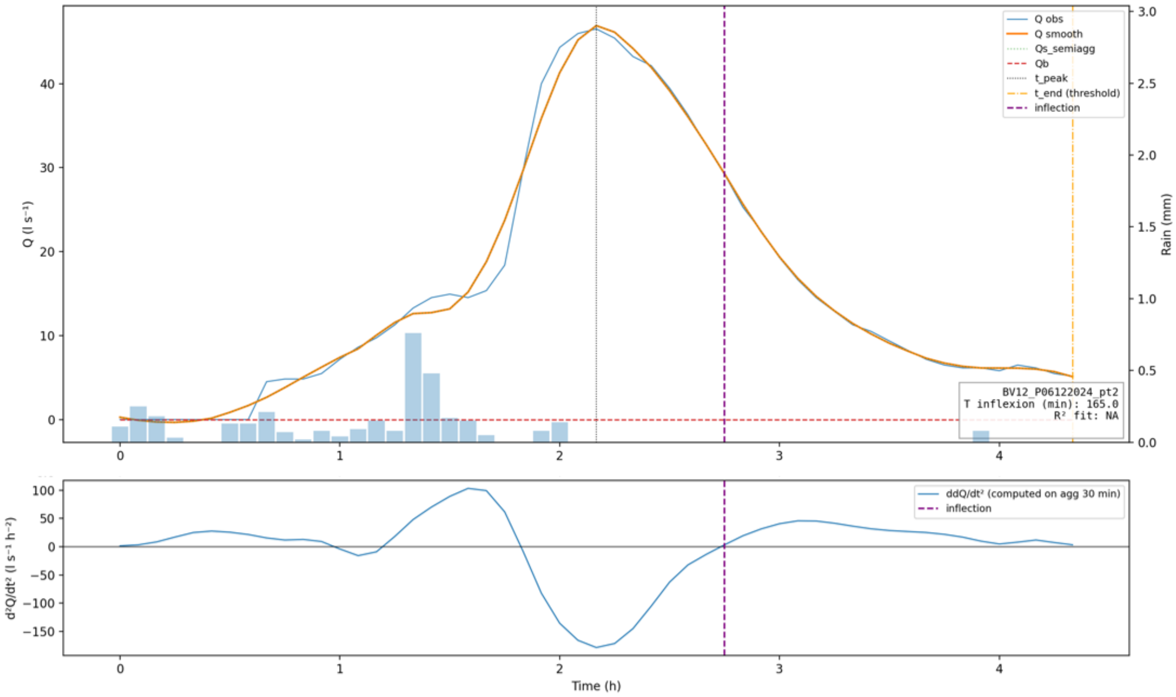

Figure 7 - Graphical procedure used to identify the inflection point on the observed hydrograph recorded on 6 December 2024 in the 84-ha Chastre catchment. The upper panel presents the observed discharge hydrograph (blue), the smoothed hydrograph (orange), and the corresponding rainfall hyetograph (bars). The lower panel shows the second derivative of the smoothed hydrograph. The purple vertical line marks the inflection point, defined by the transition in hydrograph concavity.

- The statement that the distributed model "consistently outperforms" the lumped model should be reconsidered because the numerical improvements are very small.

We agree that the numerical improvements of the distributed approach are relatively small. We will systematically rephrase claims to present a more balanced, factual, and cautious evaluation of the model's performance relative to the lumped approach.

- The reference given as "Baert et al., 2026: currently being published" should be replaced by a published article, a citable preprint, or removed.

The referenced manuscript by Baert et al. is currently submitted and under review at the journal CATENA. We will update the in-text citation and reference list to "Baert et al. (under review)" to accurately reflect its current publication status. Should the journal's editorial policy strictly prohibit citing submitted manuscripts without an available preprint, we will readily remove the reference and adapt the corresponding paragraph.

- The manuscript requires careful language editing. There are repeated typographical errors, missing spaces, duplicated words, and awkward expressions.

We appreciate the reviewer's careful reading. We will conduct a thorough and meticulous proofreading of the entire manuscript to correct any remaining typographical errors, missing spaces, duplicated words, and awkward phrasing prior to resubmission.

In conclusion, the manuscript contains a useful dataset and a technically interesting, distributed modelling exercise. However, the present form overstates model performance and transferability. A major revision should focus on reframing the contribution as a local proof-of-concept, clarifying the extent of calibration, validating or qualifying the spatial flow-path results, evaluating scale and computational requirements, and aligning all conclusions with the modest predictive performance and the single-catchment evidence.

Thank you for your review!