Reply to RC1

On https://doi.org/10.5194/egusphere-2025-5855

Reviewer comments are in black font, author comments in red font, proposed changes in the manuscript in *italic*. References to the line numbers in the revised version are indicated as rL100.

I would like to begin by thanking the authors for their research and for sharing their findings for discussion. Given the increasing frequency of intense rainfall events with significant damage potential, and the considerable efforts by federal states to produce high-resolution heavy rainfall hazard maps, advancing hydrological and hydraulic modeling remains crucial. In particular, there is an urgent need to better quantify and model how varying vegetation conditions affect runoff generation and overland flow dynamics. In this context, the authors' work is highly valuable, offering insights that are both scientifically relevant and practically applicable.

At the same time, I see room for minor improvement in the presentation of the results, with the aim of making the findings clearer and more accessible to the professional community. Before providing our detailed comments and questions, I would like to share some general critical-constructive observations:

The model setup appears to differ from the study sites described by Ries et al., as the discharge in the model is measured at a different location compared to Ries, where, to my understanding, discharge was recorded 10–20 m downstream of the plot via drainage tubes. This discrepancy could lead to distortions in the subsequent evaluation and should at least be acknowledged as a source of uncertainty. It may also partly explain the differences observed between the model results, the Feldmann data (which presumably have comparable uncertainties), and the experimental measurements.

We sincerely thank the reviewer for the positive and encouraging assessment of our work. In response, we will revise some parts of the manuscript to improve clarity and readability.

We acknowledge that differences between field measurements and model configurations inevitably introduce a degree of uncertainty. In the experiments by Ries et al. (2020), discharge was measured using a trench and drainage tube system, where water levels were recorded with a pressure-compensated piezometer and subsequently converted to discharge using a calibrated stage–discharge relationship. As reported by Ries et al. (2020), this relationship was adjusted using a correction factor derived from calibration experiments, which already accounts for part of the hydraulic effects associated with the measurement setup.

We agree that the simplification in the models may introduce additional uncertainty in the model–data comparison. Following the reviewer's suggestion, we will explicitly acknowledge this limitation in the revised manuscript and discusses as a potential source of uncertainty affecting the comparison between simulated and observed hydrographs.

Additionally, providing a rough overview of the ranges of discharge, flow velocities, and water depths across the plot would be highly useful for context and interpretation.

The range of rainfall intensity and discharge is provided in Table 2. Flow velocities and water depths were not directly measured in the field experiments by Ries et al. (2020).

The study emphasizes the influence of vegetation cover, suggesting effects not only on the roughness coefficient but also on key infiltration parameters. However, this impact may be partly inherent to the modeling approach: because the model is calibrated using the original data, where these patterns are already present, the results could reflect a degree of circular reasoning, effectively producing a self-fulfilling confirmation of the initial observations. Also, some of the figures are partly difficult to read and could be improved.

We should note that the calibration was performed just for Run2 and the validation was performed using independent rainfall events (Runs 1, 3, 4, 5, and 6) and across multiple sites. The consistent model performance across these independent conditions indicates that the representation of vegetation effects is not solely an artifact of the calibration process, but is required to reproduce the observed runoff dynamics.

We would like to clarify that our intention is not to provide independent proof that vegetation affects both surface roughness and infiltration processes, as this has already been demonstrated in previous studies. Instead, our objective is to highlight that, for accurate runoff simulation, it is essential to account for vegetation effects not only through roughness coefficients but also through infiltration-related parameters such as saturated hydraulic conductivity (Ksat). We have added rL108-rL117 in the introduction and revised section 2.3 "Parameter calibration and validation"

L10 introduce also ksat and psi to complete the picture

The revised sentence now reads:
"This study evaluates multiple methods for estimating Manning's roughness coefficient and explores the influence of vegetation on infiltration parameters, namely saturated hydraulic conductivity (Ksat) and wetting front suction (Psi), using the OpenLISEM model." (rL10-13)

L98 Original data resolution of 1min. Does this have a impact on the results? What are the uncertainties?

Temporal resolution effects in representing rapid changes in runoff dynamics and peak discharge. However, given the overall duration of the rainfall events and the relatively uniform rainfall intensity applied during the experiments, the 1-minute temporal resolution is considered sufficient to capture the dominant hydrological processes. See new section 4.1 "Applying a catchment model on a plot experiment"

L105 To explore the impact of roughness on overland flow -> To explore the impact of different roughness functions on overland flow… would be more appropriate?

We will modify the text accordingly to: *''To explore the impact of different roughness functions on overland flow"* (rL162)

L113 Microrelief / Microtopography should better fit to roughness coefficients instead of depressions which are mainly connected to retention effects

We will change it to microtopography. See rL171 and rL173

L115-120 The differences between original setup and model needs to be addressed somehow. At least that this leads to uncertainties of the results. Could be also an explanation for the differences to the results of Feldmann?

Ans: We agree that there are differences between the original experimental setup and the model representation. It is not possible to fully replicate all aspects of the experimental setup in a catchment model. For example, the model does not explicitly account for small-scale heterogeneities such as the spatial distribution of vegetation or the presence of preferential flow paths. These simplifications may introduce additional uncertainty in the simulation results. Following the reviewer's suggestion, these points will be incorporated into the discussion a new section, "4.1 Applying a catchment model on a plot experiment" (rL435ff):

''A catchment model can only applied to a small scale field experiment with some workarounds and assumptions, leading to additional source of structural uncertainty.

First, differences between the experimental setup and the model representation introduce structural uncertainty. In the field experiments (Ries et al., 2020), runoff was collected using a trench and drainage tube system, whereas in the model, discharge is recorded at a defined outlet cell. Although adjustments were made to approximate the experimental conditions, the simplified representation of the measurement may influence the comparison between simulated and observed results.

Second, the model does not explicitly represent small-scale heterogeneities such as the spatial distribution of vegetation, microtopography, or preferential flow paths. These factors can significantly affect overland flow and infiltration processes in real field conditions, but uses a simple surface retention factor as a proxy for micro-topography.

Third, the temporal resolution of the experimental data (1 minute). While the relatively uniform rainfall intensity and duration of the events suggest that the available resolution is sufficient, higher resolution would give us more detailed information about the hydrological process.
Forth, the infiltration modelling approach (Green–Ampt) assumes simplified soil conditions, which may not fully represent complex field situations, particularly under very dry or near-saturated initial soil moisture conditions. This limitation is reflected in the reduced model

performance for extreme antecedent conditions (e.g., Runs 1 and 5). However, using any explicit infiltration model for runoff modelling is already a great improvement in comparison to the still common "effective rainfall" approach based on the curve-number approach, where the estimated infiltration is simply subtracted from the rainfall, as it allows infiltration of water that has already travelled at the surface.

Differences between modelling approaches can further contribute to variations in results. For example, the model setup used in this study and the framework proposed by Feldmann et al. (2023) rely on different assumptions regarding roughness parameterization, calibration strategies, and representation of hydrological processes. These methodological differences, combined with the inherent uncertainties in both models and measurements, may partly explain the observed discrepancies between the results.

L125 model data 1s. original data 1min. How does this influence the results?

The model simulations were conducted with a temporal resolution of 1s to ensure numerical stability. However, for comparison with the observed data, the simulated results were aggregated to a 1min temporal resolution, consistent with the original measurement interval. This is discussed in the new section, "4.1 Applying a catchment model on a plot experiment" (rL435ff)

L126 … kinematic wave is used together …

we will modify it. rL185

L128 Porosity and initial soil moisture was estimated or measurement data? Same for all model runs?

Both porosity and initial soil moisture were derived from field measurements. The initial soil moisture was measured prior to each rainfall experiment and therefore varies between model runs, reflecting the actual antecedent conditions for each run.

In contrast, porosity was measured once for each site and assumed to remain constant across all model runs at that location. This approach is based on the assumption that soil porosity does not change significantly over the short duration of the experimental campaign.

L131 What are the reasonable ranges?

The parameter ranges used in this study are Ksat = 5–100 mm h^{-1} and Psi = 0–50 cm. These ranges have been included in the manuscript in rL284-287.

L142 Surface roughness functions -> maybe "Flow resistance parameterisation" is better

we will modify it. (rL202)

L146 complicated sentence, maybe: In this study, two depth-independent and five depth-dependent roughness functions were implemented in OpenLISEM to assess the impact of the investigated approaches on vegetation modelling.

We will modify this to:

OpenLISEM uses the Manning's equation to calculate flow resistance with a constant Manning's n, to be provided by the user. We have extended the code of OpenLISEM (see https://github.com/philippkraft/openlisem/tree/modified_manning_console) with five depth-dependent roughness functions (Sections 2.2.3-2.2.7) and compare the results with two depth-independent Manning's n estimations (Section 2.2.1 and 2.2.2). (rL205-208)

L150 Chow´s method could be changed to method with constant coefficients as it´s more like Manning´s method and chow provided reference tables / guidelines.

Modified. (rL211)

L153 … it does not consider the effect of water depth in presence of vegetation. -> … it does not consider the dynamic effect of water depth on the roughness coefficient in presence of vegetation.

Modified. (rL214)

L155 add that the same study site and experiments have been used by Feldmann

Added. (rL216)

L157 They estimate surface roughness -> roughness coefficients maybe more suitable?

Modified. (rL218)

L169 in equation (1): $5 \cdot h_{veg}$

Modified. (rL 231)

L185 1-2 sentences about this approach would help for understanding

Explained as: *''Following Feldmann et al. (2023), Manning's roughness coefficient is parameterized as a reciprocal exponential function of water depth, enabling systematic exploration of different roughness–depth relationships. They selected a reciprocal formulation to ensure a more balanced sampling of roughness functions across the parameter space, thereby preventing the underrepresentation of low roughness values.* (rL245-250)

L195 Only Hinsberger et al.; not in Oberle et al. 2021

Modified. (rL 259)

L208 Please insert the unknown sensitive parameters

Changed the term "unmeasured parameters" to "unknown parameters" in this context, referring specifically to saturated hydraulic conductivity (Ksat) and wetting front suction (Psi).

L211 It remains unclear on what basis the values from Chow were selected. Average? Max or Min?

The Manning's roughness coefficients based on Chow (1959) were selected according to the vegetation characteristics of each site, particularly the percentage of vegetation cover. Higher vegetation cover corresponds to higher surface resistance and therefore higher Manning's n values, while lower vegetation cover is associated with lower roughness values.

Accordingly, for sites with dense vegetation (e.g., close to 100% cover), higher values within the range reported by Chow (1959) were assigned, whereas for sites with sparse or no vegetation, lower values were used. (rL278f)

L222 (NSE) instead of L224

Modified. (rL288)

L232 MSE -> NSE

modified. (rL302)

L248 Calibration of different roughness methods -> Wouldn´t be something like "Calibration of ksat and psi for different roughness methods" be more suitable?

modified. (rL324)

L265 Lag for falling hydrograph between observed data and model maybe due to differences between model and original setup?

Explained in section 4.1 ''Applying a catchment model on plot experiments" (see above).

L279 add (ksat) (psi)

Inserted. (rL354)

L295 The font size appears to be inconsistent within the table.

Checked and corrected.

L337 line break incorrect

Corrected (rL403)

L346 What is with loc 13, run 6. Differences look significant in contrast to statement in the text.

Changed ''For these runs" to ''For these runs except Run6 in Loc13". (rL413)

L351 Isn´t this already evident from the raw data?

We agree that this trend is already visible in the raw data. However, the purpose of our modelling approach is to demonstrate that this pattern can also be consistently reproduced by a physically based hydrological model. This highlights that, for improving hydrological model performance, it is important to account for vegetation effects not only through Manning's roughness coefficient but also through infiltration-related parameters.

L357-361 From my understanding, Run 5 was carried out on the same day as Runs 2, 3, and 4. Why do the initial soil moisture conditions differ, for example, from those of Run 4?

Yes, Run 5 was conducted after Runs 2, 3, and 4, and in addition it has the most duration of rainfall, therefore the soil was already significantly wetted by the preceding rainfall events. However, after checking the initial soil moisture values given be Ries et al, the differences between run 4 and 5 are indeed nearly not existent. The reason for overall worse results of run 5 are not clear. We have deleted this sentence (rL 427): "*This substantial variation may instead be attributed to the initial soil moisture conditions, which are discussed in Section 5.2* (now 4.3)." The discussion in section 4.3 is changed to: "*Run 1 began with unusually dry soil, while Run 5 followed a sequence of moderate rainfall events (Runs 2, 3, and 4), although the measured soil moisture does not differ substantially between run 4, 5 and 6*" (rL524)

L384 emergent

Corrected (rL471).

L384 also this leads to constant manning values for some of the eq?

Yes, under conditions where the flow depth remains below the vegetation height, the Linear method can result in effectively constant Manning's n values.

L405 Again also possible discrepancies between setup of your model, Feldmann model and original experiments could lead to differences.

we explained it in the new section 4.1 ''Applying a catchment model on a plot experiment'.

Reply to RC2

General:

This manuscript provides valuable insight into the importance and sensitivity of roughness parameterization and infiltration values commonly applied in hydrological modelling and engineering decision-making. The authors employ appropriate methods to address their research objectives; however, there is room for improvement in the justification and explanation of certain methodological choices, as well as in the presentation of the results. The use of previously published data to implement the modelling framework is appropriate and well justified. I strongly recommend expanding the description of the modelling approach and the application of the referenced data in Section 2, as detailed in the specific comments. In addition, strengthening the introduction and better aligning the overall storyline with the discussion would further enhance the scientific contribution and overall impact of the study.

Thank you for the feedback, we will address these issues.

Specific comments

Lines 21-32, it makes your case stronger if you link your statements to more or general (important) references (as applied after line 32). 1A: I.e. In line 25, you cite Zhang et al. 2018, I assume there are much more (important, or general) references underlining this statement. Please check/read and especially references therein or there to:

Busari et al. 2016 https://doi.org/10.1016/j.jher.2016.02.003

Crompton et al. 2020 https://doi.org/10.1029/2020WR027194

Jackson and Klaus 2025 https://doi.org/10.1029/2025WR040753

Bachmair et al. 2012 https://doi.org/10.1029/2011WR011196

Crompton et al. 2025 https://doi.org/10.1029/2024WR037176

Thank you for suggesting these references. We have added them to the Introduction and further discussed the importance and findings of several of these studies (rL 48-66):

> *Mügler et al. (2011) compared four roughness formulations for simulating overland flow and tracer transport under rainfall simulation experiments at the plot scale. They evaluated Darcy–Weisbach, Lawrence, constant Manning, and depth-dependent Manning formulations using Saint-Venant equation. The best performance was obtained with a Manning-type model using a water-depth-dependent roughness coefficient, however, models with constant roughness coefficients underestimated high flow velocities and failed to reproduce tracer transport adequately.*
>
> *The study by Crompton et al. (2020) investigated how the selection of different roughness schemes affects hydrological predictions in modeling shallow overland flow on hillslopes with patchy vegetation. The authors propose a kinematic framework that enables the calibration and comparison of these formulations under a common flow condition. Using numerical simulations of overland flow on synthetic hillslopes with patchy vegetation, they express different roughness relationships within a unified velocity–depth formulation, cylinder array; Darcy-Weisbach; Manning; transitional and laminar. The roughness parameter is then calibrated so that flow outputs (velocity and depth) at a reference point, such as the hillslope outlet, are consistent across all models.*

In this way, any remaining differences are attributed solely to the underlying physics of the roughness formulations, allowing for a fair comparison. Their results show that, in modeling shallow overland flow, the choice of roughness scheme is less important than the correct calibration of roughness parameters. Crompton et al. (2025) also used discharge and velocity data from 112 rainfall simulation experiments to evaluate four commonly used roughness schemes by calibrating runoff model. They incorporated both discharge and velocity into the calibration process to improve the selection of an appropriate roughness formulation. Among the tested approaches, a transitional flow equation provided the best overall performance. The calibrated roughness coefficients were found to vary with surface characteristics, with litter cover emerging as the dominant control.

B: I.e. 2, Line 26. There are much more (important, or general) references underlining this statement, please check/read and especially references therein or there to (as you already do in the discussion!):

Beven and Germann, 1982 https://doi.org/10.1029/WR018i005p01311

Cerda, 1996. https://doi.org/10.1016/0016-7061(95)00062-3

Liu et al. 2025 https://doi.org/10.1016/j.jhydrol.2024.132465

Thompson et al. 2010 https://doi.org/10.1029/2009JG001134

Zwartendijk et al. 2017 https://doi.org/10.1016/j.agee.2017.01.002

We added the suggested references to the introduction (rL 108-118):

For example Liu et al. (2025) studied how different forest restoration types on the Loess Plateau affect soil water retention and infiltration. They found that natural secondary forests performed much better than planted forests by improving soil organic matter, porosity, and infiltration capacity. The authors showed that vegetation restoration improves hydrological processes mainly by changing soil properties. Thompson et al. (2010) analyzed how vegetation affects soil infiltration across different climates, from deserts to humid tropical regions, using field data and a meta-analysis of nearly 50 ecosystems. They concluded that in dry ecosystems, infiltration capacity increases strongly with vegetation biomass, while in humid regions the relationship becomes much weaker. Zwartendijk et al. (2017) by comparing abandoned agricultural land with regenerating forests in eastern Madagascar, found that soil compaction and degradation caused by repeated cultivation strongly reduce hydraulic conductivity, increasing erosion and overland flow. They also showed, secondary forest succession contributes to the restoration of soil structure through root development, litter deposition, and increased biological activity.

Line 45 "… staple area of research within hydrology", explain and/or add reference. This needs to be phrased stronger.

This sentence will be deleted. Later in the introduction we will include the following, to explain the idea behind the "staple area" statement, but with different wording (rL82-85):

These studies highlight evaluating the equations to calculate overland flows continues to be an active research area within hydrology because the appropriate roughness *formulation is difficult to determine under varying flow conditions (Nicosia et al., 2024; Crompton et al., 2025). Overland flow can shift between laminar, turbulent, and transitional regimes over short spatial scales depending on surface cover and slope (Nicosia et al., 2024).*

Lines 53-56. The coherence between the sentences is not clear. What is examined in this study and what is the additional value after the previous mentioned studies? Which findings highlight the need for further evaluation?

We will rephrase this part (rL 78-81):

Luhar and Nepf (2013) investigated the effect of vegetation distribution on channel velocity, presenting physically based models that link drag generated by vegetation at the blade and patch scale to hydraulic resistance. They have proposed an approach to calculate roughness coefficients in different conditions of water depth and found resistance depends on the blockage factor, which represents the fraction of the channel cross-section occupied by vegetation. However, their roughness equation was proposed for channels, we assess the applicability of their approach to calculate roughness coefficient in overland flow in our study.

Lines 57-59. You're totally right. You may consider re-structuring your introduction in line with previous comments on references. Try to synchronize your introduction and discussion such that your discussion finishes the line from introduction, to methods, and your results to the discussion. Try to embed the references from the discussion into the introduction.

Hydrodynamic of surface flow and infiltration effects are usually researched separately – our first approach was also completely focused on the roughness formulation. However, during our computing experiments, we found out that both needs to be seen together. We will emphasize that approach even more in the very first paragraph of the introduction and make the "first hydrodynamic than infiltration" approach more clear with subsection headers in the introduction

Lines 57-59. A pity that, again, limited references are shown. I think it is important to show that some cases/models are available incorporating infiltration and roughness (varying per land cover/vegetation type), although examples are limited. i.e.: VanderKwaak and Loague, 2001. https://doi.org/10.1029/2000WR900272, And their more recent studies, i.e. cited in Zwartendijk et al. 2026 https://doi.org/10.1016/j.jhydrol.2026.135259 and Schwemmle et al. https://doi.org/10.5194/gmd-17-5249-2024

We will the mentioned references are added to this part of the introduction (rL 91-97):

These processes have been investigated in several studies. For example, VanderKwaak and Loague, (2001) developed a fully coupled surface and subsurface hydrological model that explicitly represents infiltration and overland flow processes. More recently, Schwemmle et al. (2024) incorporated vegetation-related hydrological processes into the process-based RoGeR model, while Zwartendijk et al. (2026) highlighted that, although many physically based models account for vegetation through infiltration and roughness parameterizations, comprehensive representations of vegetation–hydrology interactions, including dynamic vegetation effects on runoff and infiltration remain limited.

Your objectives suggest modelling, but you start section 2 (Materials and methods) with the study site and experimental set-up which is not part of your study. I strongly recommend that you start explaining your approach (to model) but the use of previously gathered data conform your objectives (perhaps mention this approach already in your introduction?).

We struggled with that decision during the manuscript writing. While the model itself can be described without the experimental setup, the calibration / validation design is highly depending on the details of the experimental setup. For that reason we decided to start with the experimental description to avoid the interruption of the modelling description. We will add a start paragraph to the whole methodology section, before 2.1, to explain the composition of our study better (rL133-rL144):

> *The effect of stormflow events is difficult to observe in an undisturbed environment, as the events do not occur predictable. Highly artificial lab experiments using flumes and vegetation proxies as in the studies by Oberle et al. (2021), Luhar and Nepf (2013) and others are important to understand the flow regime, but the findings are difficult to transfer to the heterogeneity of naturally developed vegetation and soils. Sprinkling experiments can bridge the gap between lab experiment and field observation. Conducting sprinkler experiments is resource intensive and in many cases the setup between experiments is not directly comparable or the results are not fully published. Ries et al. (2020) conducted 132 sprinkling experiments on natural hillslopes at 23 sites with different soil types and land use in Baden-Württemberg (Germany), one of the most extensive datasets accessible in southwest Germany. Our modelling study builds on this dataset to test how a numerical model with a multitude of roughness approaches deals with vegetation effects on surface runoff in a natural environment.*
>
> *Section 2.1 introduces the setup and terms of the Reis et al experiments, section 2.2 presents the chosen numerical model including the extensions implemented for this study and section 2.3 the extensive calibration and validation system.*

Line 111, "The elevation model is generated". How? Lidar? Measured by...? Third-party data? Accuracy?

We revised the section for more clarity, see below

Line 113, what is the definition of a homogeneous surface?

We revised the section for more clarity, see below

Line 113-115, perhaps cite other research applying the same simplification? Or explain why.

We revised the section (rL168):

> *Artificial elevation models are generated with a cell size of 0.5× 0.5 meter, providing information about the gradient of each specific location in the database (Table 1). Since the surface of the experimental plots is homogeneous, spatial variability can be ignored in the model, the generated DEM does not include micro-topography. So, all surface parameters that may influence overland flow are represented within Manning's coefficient. This simplification does not affect our evaluation of different roughness approaches, as the conditions are consistent across all roughness methods. Therefore, the effect of micro-topography was assumed to be inherently reflected in the roughness values.*

Lines 116-120. You mention that Ries et al. (2020) measured flow at three locations within the trench, but I found this part difficult to follow. The explanation does not clearly correspond to the photograph shown in Figure 1b, making it hard to place these measurement points spatially. I recommend indicating the measurement locations explicitly in Figures 1a and 1b (e.g., using markers or annotations) to improve clarity. The same applies to the description of the trench geometry and the "one-third" section, which is not clearly identifiable from Figure 1b alone.

We added indicators to Figure 1b to point to the measurement locations and rephrase and shorten the sentence.

Lines 127-135. "are not typically measured in the field". Why not? If in general, add references. If our co-scientists of Ries et al. 2020 did not measured it (i.e. because it was not their scope), just mention that it's not measured or available.

Changing to: "*have not been provided with the dataset*" (rL 188)

Line 133 "calibrated to best match observed runoff data", is this explained elsewhere? Perhaps insert link to sub-section.

added: "*(see section 2.3)*" (rL191)

Line 133, "It should be considered that, …", do you mean "It should be kept in mind that,…"?

replaced as suggested (rL191)

Lines 143-157, a part is repetitive (introduction?). Please elaborate on the two depth and five depth-independent roughness functions, it is not clear why you introduce 2 methods but show 7 options in the fallowing paragraphs. Make sure you introduce 2.2.1 – 2.2.7. In its current form it is not clear why you introduce these different concepts. Which did you applied?

We have applied all functions and compared the results. We will clarify section 2.2 (rL 204-207):

> *OpenLISEM uses the Manning's equation to calculate flow resistance with a constant Manning's n, to be provided by the user. We have extended the code of OpenLISEM (see https://github.com/philippkraft/openlisem/tree/modified_manning_console) with five depth-dependent roughness functions (Sections 2.2.3-2.2.7) and compare the results with two depth-independent Manning's n estimations (Section 2.2.1 and 2.2.2).*

Lines 208-209, as a reader I had to check which 6 runs you're aiming for. Please consider restructuring the methodology. Also, which of the runs did you selected and why? I don't think line 218-221 is the right location to mention this.

We reformulate the beginning of section 2.3 with a reference to table 2 which includes the "runs". To make the selection of the run more prominent we are moving the explanation, which run is selected as the calibration run to the beginning of this section. The beginning of the section reads than as (rL 270):

> *To compare the different roughness equations, we calibrate the unknown sensitive parameters of the model using one of the six rainfall experiments (Run) per site, (cf. Table 2), and validate the parameters using the remaining rainfall experiments. For most locations, run 2 (100-year return period event, prewetted soil) was selected as the calibration event. Runs 1, 3, 4, 5, and 6 were selected as validation experiments. The reason for choosing Run 2 for calibration was that the soil moisture conditions during this test were neither excessively dry nor fully saturated. Since Run 2 in site 14 does not have*

any runoff, Run 6 is selected for calibration. Site 1 has a different order of experiments, so we are using Run 4 as the calibration run.

You created a very nice flow chart (of your methodology?), figure 2. Consider some minor editions conform previous comments and show this early in the methodology! I think it improves the readability of the methods section majorly! 😊

Thank you, the flow chart was very important for our study. However, we feel that for the understanding of the flow chart all information above is needed – the setup of the experiments, the different flow resistance functions and the reasoning to calibrate Ksat and Psi. We would therefore leave the flow chart at the edge between methods and results, as a summary of the complex model setup for easy reference, while reading the result section

Table 3. You may increase the readability to use ranges and just the > and < symbols, not mentioning NSE and pBias in the table as that is given in the header already. So: >0.75 for good, 0.36 – 0.75 for qualified, and <0.36 for not-qualified.

Done as suggested. To simplify the table further, we will show the ranges for the absolute pBias.

Table 1: Criteria of NSE value (Motovilov et al., 1999) and pBias.

NSE value	Abs pBias value	Interpretation
> 0.75	< 10%	Good
0.36 - 0.75	10%-50%	Qualified
< 0.36	> 50%	Not-Qualified

Line 240-241 consider a reference or longer explanation.

We insert two references: (Shahrban et al., 2018, https://doi.org/10.1080/02626667.2018.1487560 ; Lin et al., 2024, https://doi.org/10.1016/j.jhydrol.2024.131249) (rL311)

The figure (4) is difficult to interpret in its current form. In the text, negative NSE values are discussed, but these are not visible in the figure. In addition, the red bars are described as indicating model failure, yet they appear to correspond to NSE = 1, which seems inconsistent with the definition and discussion of NSE. I recommend reconsidering the figure type or its visual encoding to clearly distinguish positive, zero, and negative NSE values and to ensure consistency between the figure and the accompanying text.

We labelled negative NSE values as “failure” but missed the explanation in the text. We changed the red bar-looking background with a red cross at the x-axis and changed the caption to “the red crosses indicate negative NSE values as model failure”.

Table 4 and related text, perhaps I missed it, but how realistic are the calibrated parameters? Does it match with field or literature values?

Both values can have a broad range depending on bulk density and vegetation and were not measured in the field. We have used ROSETTA3 (Zhang and Schaap, 2017: https://dsiweb.cse.msu.edu/rosetta/) to estimate saturated conductivity from the given soil texture and bulk density. While the values fit for 8 locations, in most cases lies the calibrated conductivity more than 2 standard deviations above the values given by ROSETTA. We extended the Table 4 with the estimated values and add to the discussion (rL506-rL516):

> *By considering the soil type at each location and comparing the calibrated Ksat and Psi values with the estimates provided by Rawls et al.(1983), it becomes evident that the calibrated values are consistently higher across all sites. The only exceptions are Sites 8 and 9, which have sandy loam soils; in these cases, the calibrated values closely match those estimated by Rawls et al. (1983). Using the ROSETTA3 tool for soil property estimation (Zhang and Schaap, 2017) using the measured soil texture and bulk density, 8 locations have calibrated Ksat values inside of two standar deviations of the ROSETTA results, while 14 locations have calibrated values exceeding two times standard deviation of the ROSETTA values. Since this discrepancy occurs across all roughness function methods, it suggests that the difference is not related to the choice of roughness method. Rather, it highlights a potential limitation of applying simplified, texture-based infiltration parameters to complex field conditions. Factors such as macropores or root development may influence soil hydraulic behavior and lead to higher infiltration rates than those predicted using generalized soil texture classifications (Beven and Germann, 1982).*

Lines 284-286, did you used a statistical comparison method like Anova or? If yes, not mentioned in the methods?

We will add ANOVA to the method section – the test of difference was added late to the manuscript. Addition to Section 2.3 (rL292-296):

> *An ANOVA test was performed to evaluate whether significant differences existed in Ksat and Psi values among the different roughness methods. Model validation results from all sites and experimental runs were compared among the roughness estimation methods to evaluate their overall performance. The effect of vegetation on infiltration was assessed by comparing the calibrated Ksat values between paired sites. In addition, the influence of initial soil moisture on runoff modeling was examined by comparing NSE values across all runs at each site.*

Lines 290-294, you may consider to move these lines to the Discussion section. You may consider to add references to support your claims.

We have moved this part to the discussion section 4.2 (rL512-516)

Effect of initial condition and pre-event soil moisture. Can this be explained as temporarily hydrophobicity of the soil for example? I think it's recommended to elaborate on this in the discussion as initial condition of soil moisture has a large impact on your (and other research's) outcomes.

We would hypothesize that the effect is induced by temporary hydrophobicity – however, we would not like to speculate about this question. The overall role of antecedent soil moisture is discussed in section 4.3.

Lines 329-330, Please elaborate on this as its purpose and contribution to this paragraph is not clear to me as its not described in the methodology section. Are you comparing the individual sites or the separate model outcomes? I suggest to extend the methodology so its clear to the reader what kind of comparisons you made (and using with methods are used for these comparisons).

We added a subsection to the methodology (rL 317):

> *2.5 Comparison of paired sites*

The dataset by Ries et al (2020), uses partly a paired site approach (Table 1). The paired sites are in close spatial proximity, have similar soils but different vegetation covers. We are using these site pairs to compare the results with and without vegetation to account for vegetation effects on model performance. Due to the great efforts needed to conduct rainfall simulation experiments, the dataset of Reis et al (2020) is one of the most comprehensive collections with repeated methodology, but in effect, only six paired sites exist.

Figure 8. Are you presenting calibrated or measured Ksat values? Did you determined the vegetation cover yourself or are the values determined by Ries et al. (2020)?

We will extend the figure caption as: *Figure 8: Calibrated saturated hydraulic conductivity values (Table 4) for various vegetation coverages (Table 1) at the paired sites. The images are from Ries et al. (2020). There is no picture for Site 13 available.*

Lines 385-387 add reference, idem for 387-389.

Reference to Luhar and Nepf are included now (rL474, 476)

Lines 419-423, add references.

Done (rL520)

Lines 453-459. Please consider incorporating references from studies conducted in different climatic and geological settings to place the results in a broader context. Furthermore, I suggest replacing the phrase “validate previous research” with “are in line with studies elsewhere”, which is more appropriate and nuanced given the scope of the analysis.

To shorten the already very long article, we have just removed the sentence, as the following sentence includes the same message in more detail and references already additional literature (rL555)

Lines, 471-473, in line with remark/comment #20, please elaborate on the estimation/range on the calibrated parameters (Ksat)? Does it match with field or literature values? Which model result not only in the most realistic outcome, but also in the most realistic values for psi and Ksat values? Consider keeping this paragraph descriptive and analytical, with final conclusions presented in the Conclusion.

As the results are quite homogeneous and no clear “winner” can be found, we would like to leave the discussion about Ksat realism in section 4.2.

Technical corrections and possible editions

Line 15 (and elsewhere), you may consider changing Ksat into Ksat

Changed, we have not highlighted these formatting changes throughout the manusscript

Line 44, add additional reference.

We added:(Corenblit et al., 2007) https://doi.org/10.1016/j.earscirev.2007.05.004, (Köhler and Lewandowski, 2026): https://doi.org/10.1016/j.jhydrol.2025.134882, (Gurnell, 2014): https://doi.org/10.1002/esp.3397 (rL 67-70)

Line 75, end with a .

Changed (rL 125)

Line 231, MSE should be NSE?

Changed (rL 303)

Figure 3, in the current pdf version the figure is (too) small, make sure in the final version that it is readable!

We redesigned the figure to increase readability. We checked the colors to be colorblind friendly and made the figure printable in black and white

Figure 5. To increase the readability, you may consider to use the same layout as figure 4 (after editions).

Done

Line 370, multiple spaces between “lower” and “NSE values”.

Done (rL 459)

Line 475, I suggest to rephrase this to: “Our study evaluates seven roughness estimation methods and their impact to overland flow modelling using OpenLISEM”.

Done (rL576)

Line 476, I suggest to replace this sentence by one summarizing your applied methodology.

We have extended the sentence to complete the picture of our applied methodology (rL577)

Line 491, be specific, what is “this” model, the OpenLISEM?

Added OpenLISEM (rL598).

Reply to RC3

The study addresses an important topic—integrating depth-dependent roughness and vegetation-induced infiltration changes into overland flow modeling. However, several major and minor issues need to be addressed before publication.

Thank you, your review will strengthen our contribution and remove a number of misleading formulations.

Specific comments

Methodological clarity and justification (Lines 85–100, 125–135)

The study uses 132 artificial rainfall experiments from 22 sites (excluding Site 10). However, the selection of calibration runs (Run 2 for most sites, except Sites 1 and 14) is insufficiently justified. Why is Run 2 considered "neither excessively dry nor fully saturated"? Provide quantitative criteria (e.g., measured soil moisture ranges). Additionally, the handling of zero-runoff events (Sites 12, 16, Run 1) by excluding them from NSE calculation is problematic. The authors should discuss how this affects the representativeness of model performance.

We will add measured antecedent soil moisture for each run and each site as a supplement for reference.

If no runoff occurs, the measured runoff value is 0 for each timestep, and therefore is the NSE (a transformation of the variance normalized mean square error) in this case +∞.

$$NSE = 1 - \frac{\sum(m_i - o_i)^2}{\sum(o_i - \bar{o})^2}$$

As we do not extrapolate our findings or assume a representative selection of the sites in the original dataset, we do not see any problems from excluding the events without runoff from the analysis.

Parameter calibration strategy (Lines 205–220, Table 4)

The calibration uses 5,000 simulations per model (880,000 total) with Latin hypercube sampling. However, the ranges for Ksat (5–100 mm/hr) and Psi (0–50 cm) are very broad. Why were these ranges chosen? Are they based on prior studies or site-specific measurements? Moreover, the conclusion that "differences between roughness methods are generally not significant for Ksat ($p=0.94$)" is misleading because the calibration may not be sensitive enough to distinguish them. A sensitivity analysis should be added.

Both values depend on the distribution of pore sizes, which are heavily dependent on the root system. The presence of macropores is changing seasonal in agricultural soils, depending on crop growth stage and field operations. As these values are not measured beforehand, we follow the approach suggested by Beven (1992, https://doi.org/10.1002/hyp.3360060305), to use broad, or even unphysical parameter spaces. However, at least for Ksat values, the range fits values from the ROSETTA3 method with varying bulk density values and our own field experience.

Complex systems, like the OpenLISEM model, show often interdependencies of processes, even where they should not exist. In principle, one would not expect any effect of the surface flow resistance on the calibrated saturated conductivity, the ought to be completely independent. The result, that with a certainty of 0.94 the variation is by pure chance and not induced by the surface roughness method shows, that the processes are as independent as they should be. We see a full fledge sensitivity analysis like FAST as too complex for a minor aspect of the study.

Handling of Fu's method failure (Lines 255–265, 365–375)

The authors report that Fu's method completely failed at Sites 5, 8, 15, 20, and 23 (negative NSE, pBias = -100). They attribute this to lack of vegetation cover at four of these sites. However, Site 20 (corn harvested, 0% cover) and Site 23 (corn harvested, 0% cover) are both unvegetated, but Fu's method performed differently? Please clarify. Also, the explanation that Fu's equation behaves differently when Cp < 0.74 is interesting but not quantitatively linked to the observed failures. Provide a figure or table showing Cp values for all sites.

First, we will explain Fu's method in more detail in section 2.2.7 (rL 263): *"Where Cv is the plant basal cover inside of a flume. The coverage varied between 1.25% to 30% in the lab experiments by Fu et al (2019). For the transition of the equation to surface runoff, we are using the value "plant coverage" from the Ries et al (2020) dataset, but with the uncertainty, that the basal coverage is usually lower than the total coverage. The parameters a and b vary with vegetation type. In this study, the values of a and b parameters obtained by Feldmann et al., (2023) are used."*

Second, we will clarify the result section about Fu's method (rL 331-333):

"However, while using Fu's method, the calibration failed (NSE<0) at the sites without vegetation (5, 15, 20, 23) and at the site with scarce vegetation (8, 15% plant coverage), cf. Table 1. The other methods performed well at these sites."

Third, we will clarify the discussion and omit unnecessary details. The original discussion is partly based on a formal curve analysis, that is not shown. To simplify the message we leave this result out. The discussion on Fu's method reads now (rL461-463):

The method fails at the unvegetated site (5, 8, 15, 20 and 23) and at the sparsely vegetated site 8. Fu's function should be used only for patchy vegetated flumes, as intended by the authors and not for sheet flow regimes.

Vegetation effects on Ksat (Lines 335–350, Figure 8)

The paired-site comparison (e.g., Sites 15 vs. 16) shows higher Ksat with more vegetation cover. This is a key finding. However, the authors do not statistically test whether the differences in Ksat between paired sites are significant (e.g., t-test or ANOVA). Additionally, the role of soil type should be disentangled from vegetation effects—some paired sites have different soil textures (e.g., Sites 8 and 9: sandy loam vs. sandy loam? Actually Table 1 shows both as sandy loam, but Psi differs greatly). Please discuss.

The Ries et al 2020 dataset is one of the most comprehensive rainfall simulation datasets publicly available. Nonetheless, only six pairs exist and only five with comparable results. What our results demonstrate here is a strong effect of vegetation on soil properties, that is neglected in the current modelling practices, as Gao et al write in their opinion article (and provide numerous references for their claim). However, the soil type has of course also an influence on the soil hydraulic properties and differs between the pairs. A statistical test for differences would need extensive normalization on an already scarce dataset, which we would like to avoid here.

The Psi values are not measured, they are chosen in the calibration process. But, depending on the macropore content, the Psi value can differ widely for the same soil type. Site 8 has very young maize plants, shortly after field operations and site 9 is a pasture, with perennial root framework and less frequent operations. We extend the discussion section 4.2 with (rL 523):

While Ksat is easily measurable, the effect is also true for the Psi parameter and can hence lead to widely differing chosen parameter ranges for similar soils with different vegetation or management paractices. A good example is the site pair 8/9.

Initial soil moisture issues (Lines 315–330, 410–425)

The poor performance for Run 1 (dry) and Run 5 (saturated) is attributed to antecedent moisture conditions. The authors suggest that the Green-Ampt model assumes initially dry soil, leading to unrealistic estimates under wet conditions. This is a critical limitation. The authors should quantify the error (e.g., RMSE or bias) for these runs compared to others. Also, consider testing a modified Green-Ampt or a different infiltration model (e.g., Richards' equation) in a follow-up sensitivity analysis.

We will create a table to show for each run, how many models are good / qualified / not qualified. With this table, the reader can directly see, that run 3, 4 and 6 are more of successful validated, than run 1 and 5.

We will extend the conclusion (rL 589):

Using the classic Green-Ampt infiltration model is already a step forward, compared to the prevalent "effective rainfall" approach in surface runoff modelling. With this study, we want to emphasize, that infiltration is a neglected and underrepresented process in surface runoff modelling that is strongly influenced by vegetation. Surface runoff modelling is resource intensive, using a full Richards equation approach would increase the already hefty model runtime by one or two magnitudes. Our study shows the great need for an integrated, robust, simple to solve method to calculate infiltration during runoff events. The method must be parameterized not only by soil type, but additionally by root length, macropore density.

Statistical analysis (Lines 290–300, 305–310)

ANOVA results are reported ($p = 0.94$ for Ksat, $p = 0.038$ for Psi, $p = 0.4$ for NSE excluding Fu/Exp). However, the authors do not report post-hoc tests (e.g., Tukey HSD) to identify which specific roughness methods differ. Also, the statement "no significant differences are observed among the remaining roughness methods" ($p=0.4$) is weak because the sample size is small (only 22 sites). Please add effect sizes (e.g., Cohen's d) or Bayesian statistics.

The ANOVA is just precaution to safeguard against unwanted cross effects. There *should* be no significant roughness-method-effect on Ksat and Psi. The sample size is small, and one should not over-interpret the results, for that reason we would rather stay away from post-hoc tests and effect sizes. The figures and table do not show any clear trand to favor one method over the other, and the differences are not significant. As a result, the effect size is just a random artifact without a stable empirical fundament.

We started into this study with the expectation to find one or several methods that changes Manning's n for different flow depth, as we learned from lab experiments, eg. by Hinsberger et al (2022). The result is, using simplest method to estimate roughness, the constant method using the Chow table is not inferior compared to the more elaborated methods. We will extend the conclusion to make this reasoning more prominent in the article.

Minor Comments

Line 45–50: The phrase “continues to be a staple area of research” is informal. Replace with “remains a fundamental topic in hydrology.”

> We have moved the sentence and rephrased it as (rL82): *These studies highlight evaluating the equations to calculate overland flows continues to be an active research area within hydrology because the appropriate roughness formulation is difficult to determine under varying flow conditions (Nicosia et al., 2024; Crompton et al., 2025). Overland flow can shift between laminar, turbulent, and transitional regimes over short spatial scales depending on surface cover and slope (Nicosia et al., 2024).*

Line 85: Section heading “2.1 Model” should be “2.2 Model” (since 2.1 is study site).

Somehow the numbering got wrong while submitting, there are more problems, after section 2.4 comes section 4… we fixed the section numbering in the new version

Line 125: “Isecond” → “1 second”.

- Fixed (rL184)

Line 165–170: Equation (1) uses Strickler coefficient k_str. How is k_str estimated from Chow (1959)? Provide a reference or formula.

Changed to (rL 231):

> *The parameter k_Str=n_Manning^(-1) is Strickler’s coefficient which is the inverse of the roughness coefficient above the vegetation using Chow (1959). Then Manning’s coefficient is calculated from Eq.1, depending on the submergence ratio.*

Line 190: Equation (4) (Kadlec’s method): The assumption that h0 = 5 × plant height is arbitrary. Justify this or test sensitivity to h0.

We rephrased the sentence (rL258):

> *As they did not provide information on the h_0 value, we assigned h_0 to be five times the plant height, as indicated by the lab experiments by Hinsberger et al. (2022) to apply the Kadlec’s method in our study.*

Line 240: Section “4 Result” → “4 Results”.

- Changed “4 Result” to “3 Results” (rL 324)

Line 255–260: The sentence “Overall, the NSE values … showed limited variation at each site” contradicts Figure 4, where some sites show NSE from 0.2 to 0.9. Please clarify.

- Reformulated as (rL 331): *At sites with high model performance, the different methods yields similar results.*

Figure 3: The hydrograph comparison is difficult to read because line colors are not distinguished in grayscale. Use different line styles or labels.

- Reproduced figure 3 with different line styles and better labeling. We checked that the figure is colorblind friendly.

Line 375–380: The authors state that “water depth remained below vegetation height” (emergent flow). How was this verified? Provide maximum water depth values for each site or run.

We verified this during the analysis of the results for simulated maximum water depth for each model run. To clarify this point, we have added a table as supplementary data reporting the maximum water depth for each run. It should be noted that these maximum values occur at the outlet cell, where runoff accumulates and water depths are greatest. In the remaining cells across the site area, the water depth is lower than the reported maximum values.

Line 465–475: The conclusion that “the dataset is not sufficient to quantify the vegetation effect on Ksat” contradicts the paired-site analysis. Either provide a quantitative estimate (e.g., % increase per 10% vegetation cover) or tone down the claim.

Reformulation (rL571): *the effect of Vegetation on Ksat in a surface runoff model is visible but could only be shown for the five pairs. This qualitative observation shows an important road for the future development of surface runoff models, but does not suffice to create a ready to use formulation to include vegetation effects on conductivity.*

References: Several references are incomplete (e.g., van Meerveld et al., 2019, Line 550 has a truncated title). Please check all references against HESS formatting guidelines.

We checked the references carefully and ensured the references to be in the HESS formatting guidelines