Reply to RC2

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

RC2 is in black font, author comment on RC2 in red font, proposed changes in the manuscript in *italic*

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:

> *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:

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:

> *Considering all of these studies highlight evaluating these equations to calculate overland flows continues to be a staple area of research 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:

> *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 add the mentioned references to this part of the introduction.

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:

> *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:

> *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 will add 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*"

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

Will add: "*(see section 2.3)*"

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

Will be replaced as suggested

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:

> *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:

> *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.

Will do, 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)

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 will change the red bar-looking background with a red cross at the x-axis and change 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. However, we will compare the values with results from the Rosetta3 pedotransfer functions in the discussion (Zhang and Schaap, 2017: https://dsiweb.cse.msu.edu/rosetta/).

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:

> *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 5.1

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 5.2.

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 will add a subsection to the methodology:

> *2.4 Comparison of paired sites*
>
> *The dataset by Reis 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 will be included

Lines 419-423, add references.

We will repeat the reference to Gao et al (2023) to emphasize that we are discussing their findings over several lines

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.

We will search for additional references for the claims and include them and apply the suggested rephrasing

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 mentioned before, we will compare our results with values from the Rosetta3 pedotransfer function package. However, the range of acceptable values for Ksat and Psi is large, all resulted values fall into the realistic range.

Technical corrections and possible editions

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

Will be changed

Line 44, add additional reference.

We will add:(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

Line 75, end with a .

Will be changed

Line 231, MSE should be NSE?

Yes, will be changed

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

We will redesign the figure to increase readability.

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

We will do this

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

Will be modified

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

Will be modified as suggested

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

After rereading, the sentence can be deleted completely

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

We will revise this sentence to be more specific.