

Reply to RC3

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

RC3 is in black font, author comment on RC3 in red font, proposed changes in the manuscript in *italic*

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 $+\infty$.

$$NSE = 1 - 1/n * (\text{sum}((m - o)^2) / (\text{sum}(o - \text{avg}(o))^2))$$

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 $C_p < 0.74$ is interesting but not quantitatively linked to the observed failures. Provide a figure or table showing C_p values for all sites.

First, we will explain Fu's method in more detail in section 2.2.7: *“Where C_v 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 α and β vary with vegetation type. In this study, the values of α and β parameters obtained by Feldmann et al., (2023) are used.”*

Second, we will clarify the result section about Fu's method:

“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:

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 Reis 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 5.1 with:

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 practices. 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:

The Green-Ampt model is old (1911). However, using this standard 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 trend 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: Considering all of these studies highlight evaluating these 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... will fix the section numbering in the new version

Line 125: “lsecond” → “1 second”.

- fixed

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

- We will change line 170 to: In this study, the parameter $k_{str} = 1/(n_{Manning_Chow})$ is Strickler's coefficient which is estimated from the $n_{Manning}$ coefficients in Chow (1959)

Line 190: Equation (4) (Kadlec's method): The assumption that $h_0 = 5 \times$ plant height is arbitrary. Justify this or test sensitivity to h_0 .

We rephrase the sentence:

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”.

- We will change “4 Result” to “3 Results”

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.

- Reformulate as: *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.

- We will reproduce figure 3 with different line styles and better labeling.

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 but not saved the water table in an overview file. We will reanalyse the model results and provide the maximum water depth for the each run at each site outside the trench as a supplemental material.

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

Reformulation: the effect of Vegetation on K_{sat} 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 ensure the references to be in the HESS formatting guidelines