Comment on hess-2021-551
Anonymous Referee #2

Referee comment on "Extraction of roughness parameters from remotely-sensed products for hydrology applications" by Charlotte Marie Emery et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-551-RC2, 2022

Review Emery et al. 2021

This paper aims to deliver a physically based approach to estimate roughness coefficients for main channels and floodplains, using readily available Earth Observation data. There are major issues with this paper, both in the proposed methodology, and the experimental design. A very major revision is needed to consider this paper for publication. I have listed my major concerns below. These are followed by more detailed comments.

General larger comments:

- The quality of English is insufficient. For example, the first sentence in the abstract is grammatically not correct, and also may confuse a reader whether “estimation of river discharge is indirectly derived from ...” by the authors as the subject of this paper (which I assumed), or whether this is a generically sensible approach. Please make sure a near native English speaker reviews the manuscript thoroughly.
- The abstract does not at all discuss the limitations of the approach. I see the possibly large amounts of noise introduced by the enormous amounts of introduced degrees of freedom, as a major obstacle for the use of this method. This has not been tested, and therefore, I would judge that simply choosing two roughness coefficients from text books may perform within the same limits of acceptability, at a much lower level of uncertainty.
- The basis for the approach is Arcement and Schneider (1989), which is not a scientific publication. Since the suggested approach is entirely dependent on this source, I cannot judge if the approach is scientifically grounded. Reference is needed to a scientific (peer reviewed) article, or that article should first be written before this publication can pass. At the very least, the original method should be more elaborately described and shown to be scientifically sound in the first place.
The general (and largest) problem I foresee with the suggested approach is the level of noise using so many degrees of freedom (6 parameters whereas most hydrological applications use 1 or maybe 2, and a very large allowed spatial distribution!) with such noisy data as soilgrids, and a simple land cover classification with lookup tables, with unknown uncertainties. The authors do not demonstrate that the amount of degrees of freedom are warranted and contain any predictive value. Moreover, the classification of the original approach also contains very large ranges. If these ranges would be explored in a sensitivity analysis, I can already predict that the level of uncertainty in outcomes Manning's coefficients will be very large. The argument that using this approach yields better values than non-calibrated models is not demonstrated because no proper benchmark experiment has been established and uncertainty of the estimation method has not been considered at all.

A strong reliance on the “IOTA2” approach is suggested but it is not clear why particularly the IOTA2 method to land cover classification should be used. Why not any other land cover method based on medium resolution optical imagery?

The DA experiment (3.1) is not clearly described. What do you expect to improve in data assimilation with your method? And how do you test this exactly? It looks like you are comparing one uniform manning roughness against 14 different manning roughnesses (i.e. many more degrees of freedom). That makes it highly logical that a DA experiment (or any calibration experiment) leads to a better fit of observed values, regardless of the method used to set a prior estimate on the roughness values. I don’t see the added value of the proposed method as one does not need such a method to impose more degrees of freedom on a 1D simulation model.

The SWOT experiment in 3.2 is also not clear. How are your observations introduced in the experiment? And what is the experiment exactly? Also here, is the better result not merely a result of the fact that you introduce more degrees of freedom? This would render the experiment invalid as it is not a fair comparison.

Conclusions: given the likely invalidity of all experiments I have seen in this paper, if this paper is considered for improvement, all conclusions will alter as well. There is no discussion on the limitations of the method and certainly the fact that one introduces many additional degrees of freedom (both due to multiple manning roughness components and due to the spatial distribution). No proof is given that introduced noise because of so many degrees of freedom is at satisfactory levels. The uncertainty must be investigated and discussed.

Detailed comments:

Abstract

- there is no information about the EO methods used to estimate the roughness in the abstract. I would add at least one sentence on this method.

Introduction
I would introduce the Manning-strickler relationship earlier (around l.40) and decide whether to use Manning \((n)\) or Strickler \((k)\) for the remainder of the paper, as \(n\) is simply \(1/k\).

- 70 applicable □ applicable.
- 104. What does the meandering ratio mean? And does it apply under all circumstances? If a natural river is at low levels, it follows the stronger meanders of the permanent bed, whilst at higher flows it will follow a shorter path, i.e. between the natural levees.

2. Method

- How does this work where soilgrids is not resrepresentative for the river bed? For instance, in smaller mountaineous streams, sol grids does not at all constitute a representative database. Even for larger alluvial streams, the river bed sediment may already be very different from the floodplain sediment where finer grain sizes will be dumped during floods. Are you now assuming these are always equal?
- 152. “There are a few locations where SoilGrids provide no data. In this case \(n_b\) is computed as the average of the three adopted values of \(n_b\) values, equal to 0.0245 s.m \(-3\). This is not clear. Which “adopted values?”. And it looks like for no data areas, you do the same as for with data values.
- 163: cross-sectionnal should be cross-sectional
- computation of \(n_1\). It is not clear where the cross-sectional profiles should come from if this method should be entirely remote sensing based. Another problem is that the suggested ratio is strongly dependent on the resolution of the profile observations AND the level of smoothing, which also seems arbitrary (it was supposed to be described in Appendix B but it is not, only examples are shown). If a surveyor measures a profile point every 1.0 meters, you’ll get a very different measure compared to when an observation is taken every 0.2 meters, which I find very problematic for a suggested generic approach.
- Computation of \(n_2\). Similar to \(n_1\), the choice of distance between sampled widths is not elaborated upon, does impact strongly (of course!) on the variability of the derivative of width in space, and it is also not clear how the authors then can justify the relationship between this sampled width and a contribution to Manning’s \(n\) (i.e. the “adopted value” in Table 6).
- Same story as with \(n_2\) and \(n_1\), there is no support for the mapping of the parameters to the land cover maps.
- Meandering coefficient. I think it makes sense that sinuosity can be assumed to have no impact on the floodplain, but please then describe why. And also here, a relationship is made between the indicator and the table by Arcement and Schneider, without any reasoning.

2.3 validation

- 264-265 “Its results are compared to the same observed data and the method is deemed validated if the estimation performance is close to the reference performance.”. This experiment is not valid for 2 reasons:
- There is no properly defined benchmark, i.e., a different method to sample roughness without calibration than presented in this paper. A valid experiment would be to generate several models with *a priori* sampled roughness coefficients, following typical lookup tables, as one would do without having the presented approach.
- Second no notion has been taken of noise in the sampled estimates. Given that there is a strong variance in all coefficients, and additional noise in the relation between the lookup tables and the presented components of \( n \), and there is no reason to suggest that any of the presented coefficients covaries, I expect that the noise in your \( n \) estimates will be very large and that the noise in river flow will hence we disproportionately large to render a method with so many degrees of freedom useful.
- 278 – 292. It is not clear how far the boundary conditions are from the evaluation station. If these are too close, then this will affect the results (e.g., backwater from downstream boundary condition). Has this been checked?
- Results of the validation are presented in the methods section

Section 3.1. The experiment with Data Assimilation is not properly described. What hypothesis is exactly tested? And what is the experimental design leading to that test?

Section 3.2. Same problem as 3.1. Experiment is not clear and I fear that the experiment is invalid.

- 15-17. It is not clear what the different labels mean. I guess “target” is the “observation”, but what does “prior VDA” and “real-time” mean?
- Fig 15: The blue and black line are nearly on top of each other, which makes the suggestion that a “47% reduction in errors” rather superfluous. My impression of the results is that there is overall no real difference (comparing all 3 locations) and the improvements over the Po could easily be explained by the fact that your method introduces more degrees of freedom.
- 449. “serenely”? Should this be “securely”?