The paper presents modeling results for infiltration into porous media with low initial saturation and addresses the representation of water fingers with an overshoot at the tip that has been discussed intensely in the literature in the last years. The focus is here on a reproduction of the finger width depending on the initial saturation and the related finger speed from experimental results from the literature. The important point is to capture the non-monotonic relationship found in the experiments. The model is a space discretized mass balance and it includes hysteretic capillary pressure saturation relations. The topic is of interest for the journal. However, I think that a lot of clarification is needed. I find it a great achievement that experiments and in particular the dependency of finger width on initial saturation could be reproduced with a model. But the paper makes the point that this is due to a novel model approach that is used here. The model has been published in previous papers and the ideas behind it are here not explained. However, as the claim is made that the key to reproduce the experimental findings is the model approach that is different from previous ones, one cannot follow the reasoning. Although it is understandable that one does not want to repeat too much previous work, enough explanation is needed to make it possible to follow the lines of argumentation.

- The biggest problem I have is that it is unclear to me, how the model (not in its concept but simply the set of equation that is solved in the end) differs from the discretized Richards equation solution with a specific hysteretic capillary pressure saturation relation. The authors stress in line 292 that the model is not a numerical scheme to solve the Richards equation. But they do not explain what the difference is and what could not be reproduced with a discretized Richards equation. Also, it is written that the choice of the grid blocks has physical interpretation, but they do not give this interpretation. It is referred to a previous paper (Vodak et al., 2022), but the main story line of a paper should be understandable without reading further papers.

  The difference between this model and a discretized Richards equation does not become clear to me. I understand that the choice of capillary pressure curve is made dependent on the grid size (or block size, without explanation of the concept it is not so clear why this should be different) and that this is somehow related to not covering an REV with the grid size. This is called scaling. But apart from the reasoning, one chooses in the end a (hysteretic) capillary pressure curve. This leads to the same system of equations that one would obtain with a discretized Richards equation with a specific choice of parameters. Eqs. (1) and (4) combined with a capillary pressure saturation function would be the same equations one would solve if one discretizes the Richards equation with a standard finite volume, two-point flux approximation scheme and explicit Euler time discretization. The saturation would be represented exactly as is outlined in lines 67-71.

  The authors use a hysteretic capillary pressure saturation function and the shape depends on the grid size. With decreasing grid size, the chosen curve gets flatter. They stress that the choice of the grid size is not trivial, but this is not further outlined. In the end, one fixes a grid cell and with this one chooses a capillary pressure saturation function. Whatever the reasoning behind this choice is, the function is fix. It would not make a difference if the function would have been chosen with a different reasoning. Finally, one solves the same system of equations as one would if one solves the Richards equation with a finite volume scheme using the same hysteretic capillary pressure function. One would also generate the same result. So the unstable infiltration and the finger width behaviour would be reproducible with the standard Richards equation using the same hysteretic capillary pressure function that has been used here. No new model is needed for this. I guess the key point is thus the hysteretic capillary pressure function and the choice of the function that depends on grid size. I assume there is more to the choice of the capillary pressure function and that one could predict is from material properties and knowledge of the flow regime. That would make a difference, because one could then say that this model is predictive, while the discretized Richards equation is not. But this is only guessing.

- It is claimed that using the geometric mean of the cell permeabilities to approximate the flux across volume interfaces is usually not done, but at this point I disagree. Often one uses the geometric or the harmonic mean of the saturated permeabilities and upstream weighting of the relative permeabilities, but to my knowledge it is as common to use geometric means of the total permeability. In textbooks on numerical solutions of two-phase flow equations all these options are usually discussed. It is interesting though, that the geometric mean is a key element to reproduce the fingers.

- The authors argue that the model converges to a new type of model if Δx goes to zero. If I understand correctly, this goes with the changing capillary pressure saturation curve, which goes from steep to flat with decreasing size Δx. As the authors write in line 278, the model should be able to reproduce the standard Richards equation behaviour. I do not see how the drainage from a fully saturated soil column towards a hydrostatic profile should be reproducible with this model as Δx goes to zero. The hydrostatic saturation profile should match the primary drainage capillary pressure curve. If the curve gets flat with decreasing Δx, one would not be able to retrieve the profile. One would get a sharp change from fully saturated to dry.

- I understand that the authors want to acknowledge the experimental findings of Bauters et al., 2000, and to highlight their achievements by calling the nonmonotonic dependency of finger width with initial saturation Bauters' paradoxon. Still, I find this wording a bit odd. A paradoxon involves a self-contradicting aspect or something that is against the intuitive expectation. This should not diminish the observations, and there is maybe not an easy explanation, but I find it hard to see a paradoxon.

- It is not clear if the model results presented here are predictive or if parameter and other adjustment was involved. In line 245 it is written that a four times lower infiltration rate than in the experiments was used. Why was this lower infiltration rate chosen? Was the match with the experiments not obtained with the same infiltration rate as in the experiments? I find this an important point, as a model needs to be predictive, meaning that one should be able to know the parameters from information about the materials or from measurements of the materials. It is an achievement to reproduce non-monotonic finger width with initial saturation, but if this was obtained with a model that needed fitting, one could argue that one would have obtained the same with a classical (hysteretic) Richards equation model by fitting the capillary pressure curve.

- The discussion on the REV in lines 42-51 is a bit long. The problem of the REV for two-phase flow problems has been discussed a lot (for example already in the papers on volume averaging of two-phase flow) and it is acknowledged that the REV for fluid content is problematic, in particular for unstable displacement. The question of an REV for pressure in porous media has also been discussed in the literature (just one example: Nordbotten et al., Water Resources Research 2008). I think this could be shortened and does not need all the citations.
- Line 56: I think that at least the papers of Lenormand et al., 1983, or Wilkinson, 1986, are here misleading. The point in these papers is not to derive alternative models to the Richards equation but to capture viscous and capillary unstable immiscible displacement. To my knowledge, they do not include gravity.
- Line 81: This new type of mathematical models of Nature sounds a bit overselling. The switch from parabolic and hyperbolic for two-phase flow problems is known for continuum models in the limit that capillary effects vanish (also for immiscible displacement in the fractional flow formulation).
- Line 237: Where does the 0.0005 come from? It is in none of the figures.
- Line 242-243: I think this is a bit simplifying. The solutions of the Richards equation without hysteresis are stable, so of course this effect is not captured.