Reviewer 2: The authors compare two random walk approaches to simulate Richard's equation. One is straightforward (called naïve) and the author is based on the fact that particles are distributed among the different capillary tubes given by the retention curve. The latter is a nice approach worth of publication.

Erwin Zehe (EZ): We sincerely thank the anonymous reviewer for his encouraging comments and helpful comments.

Reviewer 2: I am missing technical details on the applicability of these two methods in the paper in regards to the gradient of dispersion. I do not see how it is estimated. Also, an explanation of why the first method does not work. In theory, one should expect that the naïve method works for vary large number of particles? What are we missing here? Please, explain exactly how the two methods are implemented step by step.

EZ: The spatial gradient of the dispersion coefficient is estimated by means of a centered finite difference. This is straight forward, as the dispersion coefficient is in both approaches well defined in each grid box. We will better explain this point in the revised manuscript.

Based on the widespread evidence that particle tracking is suitable for simulating solute transport, we also expected the "naïve" approach to work well for large particle numbers and a suitable updating rate of the dispersion coefficient. We started initial simulations with the naïve approach with 10^4 particles and found this approach systematically too overestimates depletion of gradients and the thus vertical redistribution of water compared to the Richards model. Neither an increase of the particle number to 10^6 nor an iterative updating of the dispersion coefficient nor shorter time steps fixed the problem.

Within the naïve approach all water particles in the pore space experience the same dispersion $D(\theta(t))$. This assumption is based on the analogy to the solute transport problem, where all solute particles in a flow field experience indeed the same dispersion: they experience so to say the same "average path length", which implies that the diffusive step scales for all solutes with the same dispersion coefficient D^{solute} as follows $Z\sqrt{6 \cdot D^{solute} \cdot \Delta t}$. This is however not the case for water parcels/molecules in porous media, because diffusive flow velocities decrease with decreasing pore size. To account for this the diffusive step **cannot** scale for all particles with same maximum $D(\theta(t))$, it needs to reflect the distribution of D within the different wetted pore sizes fraction. To achieve this we subdivide the particles in a grid cell into N bins (800) and calculate D starting from the residual moisture content to the θ_r stepwise to $\theta(t)$ using a step with $\Delta\theta = (\theta(t) - \theta_r)/N$. The diffusive step for particles within bin i scales $Z\sqrt{6 \cdot D(\theta_r + i \cdot \Delta\theta) \cdot \Delta t}$, for i=1 … N. We will add a similar explanation to the revised manuscript.

Fortunately Reviewer 1 pointed out that our implementation of the Langevin Equation was not entirely correct (see our corresponding reply). Figure 1 shows a simulation with the corrected model in the full class mode (using 200 bins) and the "naive" approach, also based on the correct form of the Langevin equation. The full class model shows a nearly perfect match of the Richards solver, while the naïve approach shows the above explained deficit.
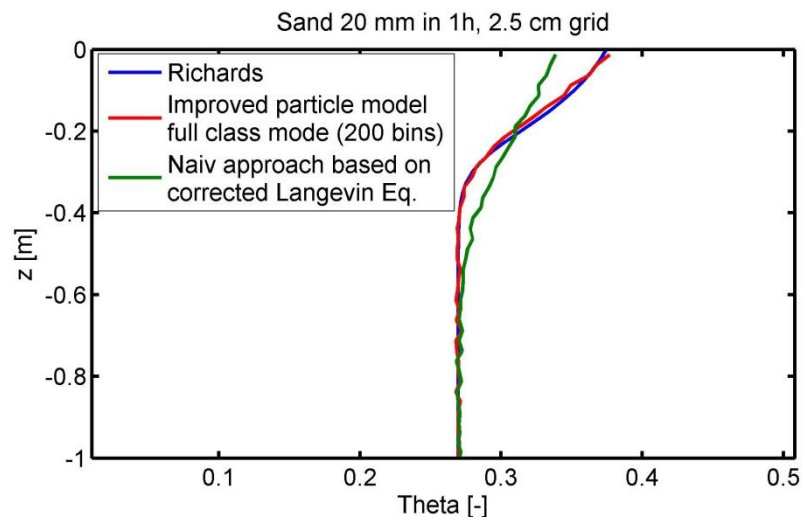
Figure 1: Comparison of the new particle model based on the corrected form of the Langevin Equation using N(0,1) random numbers in the full class mode, with the naïve model based on the corrected Langevin Equation.

We again thank Reviewer 2 for her/his insightful comments that will surely help us to improve the presentation quality of our study.

Erwin Zehe