This study by Vasconcellos et al. applies a distributed version of the “Tank Model” (DTM) to a relatively small catchment in Brazil in order to simulate soil water storage, derive a soil moisture proxy (Soil Water Index, SWI), and compare SWI to measured soil moisture and to topography-derived indexes (HAND, TWI) that could also be used as proxies of soil moisture. The goal of this approach is to show that a simple hydrological model can be applied in a distributed manner to estimate the spatial distribution of soil moisture in a catchment, which acts as a major control on hydrological and biogeochemical processes.

I think that there is an interesting idea at the origin of this work. However, in my opinion this manuscript would need considerable improvements to reach the quality required for publication in HESS, both in substance and format. I think that the methods could be carefully reconsidered and revised. At least, there could be clarifications and explanations about the choices that were made in the methods (especially on calibration and validation), and an extended discussion about the impact of these choices on the results and interpretations. The quality of the presentation of the results could easily be improved, in particular the equations (e.g. correct indexing and naming of symbols) and the figures (labels, axis names, legends…). Here are my major comments:

1. In this work it is not immediately clear why one should bother with a hydrological model which largely simplifies transport processes in the soils and comes with many sources of uncertainties (P, Q, ET, parameters) instead of using simple topographically-derived or precipitation-derived proxies of soil moisture. Why use a model if simpler metrics are satisfactory (especially when keeping table 10 in mind)?

2. Given the relatively small size of the catchment (6 ha), the assumption of spatial homogeneity of inputs / land use / soil water parameters, wouldn’t a physically-based distributed model (e.g. CATFLOW, Zehe et al., 2001) be more appropriate to simulate soil moisture? This question arises specifically after such models are mentioned in the introduction (L30-36) but not discussed further. In addition, physically-based distributed models would allow a direct and meaningful comparison of simulated vs observed soil moisture.

3. The equations could be re-written in order to avoid current ambiguities:

- Using subscripts and superscripts for the indexes (0, 1, 2…)
- Writing any explicit dependence on time t with (t), and avoiding using a subscript for this
- Replacing t-1 by $t-\Delta t$
- Correctly placing the bound variables (i, j, c) below the sums and giving the lower and upper bounds
- Avoiding using a dot symbol between bound variables and elements in the sum (this will be confused with a product)
- Correcting equations 7 and 8; Δt is missing to go from fluxes (L/T or L^3/T) to water amounts (L or L^3)

Most of my concerns are about the calibration-validation choices:

4. Would it not make much more sense to calibrate the DTM rather than transferring the parameters of the calibrated TM to DTM? In the end, the procedure would be the same, but the DTM would be used in the DREAM algorithm instead of TM. Seeing TM working better than DTM in both calibration and validation (tables 5 & 6) suggests that it should have been done

this way. As a result, the derived uncertainties are all coming from the calibration of TM, not from DTM which is the model actually used to derive SWI. Thus, the true uncertainties which will affect SWI estimations are not shown.

5. The models are calibrated to hydrographs and not soil moisture which is the target. Very good model fits to hydrographs may not guarantee very good fits to soil moisture (e.g. in a catchment dominated by groundwater responses, Loritz et al., 2017; Rodriguez et al., 2019). Is it not possible to similarly calibrate and validate DTM, but to soil moisture observations instead?
6. The visual comparisons for soil moisture are misleading. Why show only 2 locations out of 9? Is it really meaningful to compare SWI and soil moisture directly? Figure 8 suggests that their relative variations differ by orders of magnitude (especially in Fig8b), and that they may be poorly related (especially in Fig11, the step-wise behavior). The use of a correlation coefficient does not take additive and multiplicative biases into account (Legates and McCabe, 1999), so I suggest to use another performance measure. Also, the correlation coefficient does not allow a spatially-distributed comparison of the results with the observations, while it seems important for the aims of this study (see useful suggestion from first reviewer)
7. How were parameter sets chosen (L192)? I think that the use of an average parameter from all calibration events is not a standard method. In addition, no verification was done in order to check that the mean parameter set actually works well for the calibration events. In my opinion, the standard method would be to calibrate the DTM to all 5 events simultaneously, and validate to the 2 last events simultaneously. Shouldn't DTM work well for all events, to be representative of transport processes in the soils?
8. Was the Global Likelihood of Schoups and Vrugt really used? Why choose a constant homoscedastic gaussian likelihood error model, which is already available in DREAM, then?
9. How was "total uncertainty" derived? In what way does it differ from parameter uncertainty? How does it affect SWI uncertainties (seemed to be one aim of the study)?

Lastly, I encourage the authors to use comments in the code provided on GitHub and used to generate results, in order to make it more understandable.

Here my many minor comments, organized in a list for efficiency:

Comment	Lines / Table / Figure
Reference missing	L15-16
Wrong reference	L31 (wrong year, use 2017)
The language needs corrections	L75, 100, 168-170, 187-188, 201-202, 205, 214
Ambiguous or vague statement	L109-113, 151-152, 201-202
Verb missing	L18-24
Scale missing or ambiguous	Fig1b, Fig 9
Units missing	Fig1a, L127-128, Fig4
Axis names missing or ambiguous	Fig2, Fig4-5-6, Fig9
Date ticks missing	Fig2, Fig8
Captions not explicit enough	All figures
Errors in the caption	Table 6
Details missing on the methods	L85-87, L109-113, L172-173
Clarifications needed	Eq1, L94-98, 114-116, 151-152
Undefined symbol	Tab2 (ETR)
Reformulation needed	L114-116, 134, 164, 181-182, 184-186
Symbol already used	L170 (T)
Wrong symbol	Eq12 (t)
Repetitive sentences	193-196

References:

Legates, D. R., McCabe, G. J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resources Research, 35, 233-241, 1999.

Loritz, R., Hassler, S. K., Jackisch, C., Allroggen, N., van Schaik, L., Wienhöfer, J., and Zehe, E.: Picturing and modeling catchments by representative hillslopes, Hydrol. Earth Syst. Sci., 21, 1225–1249, https://doi.org/10.5194/hess-21-1225-2017, 2017.

Rodriguez, N. B., Pfister, L., Zehe, E., and Klaus, J.: Testing the truncation of travel times with StorAge Selection functions using deuterium and tritium as tracers, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-501, in review, 2019.

Zehe, E., Maurer, T., Ihringer, J., and Plate, E.: Modeling water flow and mass transport in a loess catchment, Phys. Chem. Earth Pt. B, 26, 487–507, https://doi.org/10.1016/S1464-1909(01)00041-7, 2001.