Review of **Evaluation of Soil Water Index of distributed Tank Model in a small basin with field data** by Vasconcellos et al.

The study of Vasconcellos et al. applied the Tank Model and its distributed version D-Tank to the Araponga river basin in Brazil. Here, soil moisture values were measured and the authors calculated the soil water index (SWI) of D-Tank. The authors correlated the soil moisture measurements with SWI. Afterwards, the SWI was correlated with the Topographical Wetness Index and Heigh Above Nearest Drainage. The authors conclude that SWI is a useful metric to represent soil moisture.

The study seems interesting, but I think some effort is needed to improve the manuscript. Generally, there are often sentences that seem off to me (see also my long list of minor comments), and I think the authors should go over it once more to capture all these small errors. Also in the Figures and Tables there are many minor issues as missing units or different color scales.

In addition, the authors make often statements which are hard to find back in the figures or need a bit more support. For example, based on Figure 8, the authors state that there is a satisfactory performance, but Figure 8b seems quite off in my view, so I am not sure if you can make this statement. More importantly, there are many more tensiometers and events, so why not show more (ideally even all) data? This would generalize and support the statements of the authors much more. The same applies to Figure 11, even though the correlation coefficients are reported in Tables 7 and 8, more results can easily be shown. In addition, a more critical look at this figure is probably justified, instead of just looking at the R-squared values. For example, why are the soil moisture values flattened off? And why is the SWI just changing after a soil moisture value of around 0.155 cm3/cm3? Just looking at the R-squared value gives a good indication of the linear relation, but it does not say much about, for example, a consistent bias, so I might be good to look a bit further here.
Related to this, I also think it is important that the authors assess the spatial correlation between SWI, HAND and TWI a bit more thoroughly, instead of just a linear regression. For example, normalizing them and subtracting the maps from each other will lead to insights where the values match, and where deviations start to occur. This can also easily be done for the interpolated values of soil moisture (Figure 13).

Regarding the calibration, what is actually the rational behind using the DREAM algorithm? First, I am a bit confused on how it was implemented, there is a mention of a generalized likelihood function, but in the next sentence (P9.L174) it is stated that the last 7500 samples are used to represent uncertainty. So how was the uncertainty actually represented in the end? In addition, the DREAM method is applied to each event to get parameter estimates, but in a next step, the parameter values of the different events are just averaged, which seems a bit simplistic in contrast to the DREAM algorithm. Why not give more weight to the parameters that are more likely? The added value of this method is also that it gives you uncertainty values, but afterwards, the authors do not really do anything with it in the analyzes of soil moisture indexes. So what is the point of using this method? In addition, I think there is also a strong influence of the chosen evaluation statistics, as the Nash-Sutcliffe efficiency and the RMSE have a strong bias for large values. In other words, in the event based approach in this study, when the height of the peak matches, high values for these metrics are likely to be found.
The authors also never specifically mention how they deal with the initial states, which will have a strong influence on the results. Is it correct that these are calibrated?

Lastly, I also wonder if it is really a surprise that TWI, HAND and SWI show good correlations. In the end, they are all based on the flow direction map, and especially on the event time-scale for a small

basin, I think they should show similar patterns. I am not sure about this, but I just wonder what the authors thoughts are here. Would it not be much more interesting to see if these findings still hold for a longer time-scale and/or even a larger area?

Concluding, I believe there are some substantial improvements needed. I hope the authors find my comments useful and I look forward to an improved manuscript.

Minor comments

P1.L18. Condition → conditions?
P1.L24. (Beven and Kirkby, 1979) → (Beven and Kirkby, 1979).
P2.25. Have been → has been
P2.L27. I do not think you can see the hand as model, more a corrected elevation map
P2.L65. clayey texture. → clayey texture
P2L67. Clay) → clay
P3.L75. Sentence seems odd, please rephrase.
P3.L78. At the 0.3m → at 0.3m
P3.L85. As showed → as shown
P5.L95. (Sugawara, 1995) → Sugawara, (1995)
P5.L100-101. there were selected 5 rainfall events → there were 5 rainfall events selected
P5.L111. I think this depends on the basin. Surface runoff will only occur if there is infiltration excess (so depending on the rainfall rate and infiltration capacity) or saturation excess (depending on the saturated state of the catchment).
P6.L120. I would suggest to use subscripts, it is not clear if it is H*A1 or HA1
P7.L131. soil type and use as well as → soil type as well as
P9.L162. And after transformed → and afterwards transformed
P9.L163. Please be more specific. HAND requires defining also the stream, as you have to define where the stream starts (i.e. number of upstream cells). Or do you have this mapped? And what do you mean with reclassified?
P9.L169. soil moisture distributed models → distributed soil moisture models?
P9.L167. proposed Laloy .. → proposed by Laloy …
P9.L168. Used for → used to
P9.L188. it was adopted 2-m resolution → a 2-m resolution was adopted
P10.L194. Why do you just use the average if you do such a complex calibration?
P10.L197. Should you also not look at the spatial distribution?
P13.L225. You cannot state you have a good water balance based on the NASH and RMSE values.
P15.L230. In the similar way → in a similar way?
P15.L233. Yes, the baseflow coefficient must be off, but why should this be overestimated? Also in calibration the declining limbs are off.
P16.L239. behavior comparison?
P16.L241. of the Figure → of Figure
P16.L242. How is it satisfactory? The lines in Fig. 8B are completely off.
P17.L249. I find this a bit hard to see, it is not really different compared to other values along the stream.
P18.L251-253. The legends actually say the opposite, that the maximum values for event VI is 251mm and for VII 214mm.
P18.L255. I cannot see from the figures that the values are higher at the outlet.
P18.L264. I cannot see this, can you add the location of the river in the figure?
P18.L269. the only measured → the measured
P18.L270. What do you mean with minimally coherent?

P18.L272. Sentence misses a verb.
P18.L273. What is the value of 0.7? Correlation coefficient?
P18.L274. of to the → of
P18.L275. Please also report significance levels and p-values.
P18.L280. The comparison would be easier if 12 a and b both use a continues, similar colorscale. You could also normalize the values for comparison.
P19.L282. How was this correlation analysis carried out? Does the regression use all the cell-based values of SWI and HAND? Why not show a graph of this?
P19.L285. Explained for → explained by
P19.L286. Major → stronger?
P20.L290. What do you mean with coherent?
P20.L293. for estimate → to estimate
P21.L296. How do you define the covered area?

Table 3. Are S1I and S2I the initial states? Do you calibrate these?
Table 5. Please add the units of the RMSE. What is the range of values for NASH and RMSE here, based on the uncertainty ranges?
Table 7-8. Why are there so many places left empty? What are the criteria to leave these out?
Table 10. All the values are the same for the different events, is that correct? Are I, II, and III referring to the different moments in the hydrograph?

Fig1. Please add units of elevation.
Fig4. Please add the units. I also assume the dependent variable is flow, and the independent variable is time, correct? Can you add this instead? What do you mean with uncertainty of the parameters? How is this different from the grey area?
Fig5. What is the difference with Fig.4? What happened to the uncertainty margins?
Fig9. Please make sure all figures use the same color scale. It has hard to compare like this.
Fig10. The second graphs is also labeled as Cross Section I.
Fig11. Why are the soil moisture values flattened off? How is this regression carried out?
Fig13. Please use the same color scale for comparison.