Interactive comment on "Information – based uncertainty decomposition in dual channel microwave remote sensing of soil moisture" *by* Bonan Li and Stephen P. Good

Anonymous Referee #2

Received and published: 7 December 2020

The topic of the manuscript is certainly important and interesting for HESS readers. However, the manuscript contains a large number of typos, things that should be explained in more detail and assumptions that are not discussed and could affect the results.

The authors cite (Lines 84-86) just two of the possible uncertainty sources of the retrieval model, however the method discussed in this manuscript is based on in situ measurements. Therefore, the uncertainty of in situ measurements must be taken into account, and it is not at all. In situ measurements are here considered as "ground truth". Unfortunately "ground truth" does not exist, as all measurements, they have

errors. But probably more important is the uncertainty of the spatial representativeness (satellite Tbs are representative of a spatial scale of tens of kilometer while in situ measurements are single point measurements) and depth representativeness (sensors measure at a given depth while the Tbs are representative of a different/changing depth). These effects must be mentioned and discussed and their possibles effects on the results should be analyzed.

In addition, why using 9km Tbs instead of the original Tbs in the 36 km grid which is closer to the instrument resolution ($\sim$ 50 km). The SM dataset that is provided in a grid with 9-km sampling has been obtained using a Backus-Gilbert interpolation. Surprisingly this is not mentioned at all in the manuscript. How could this choice affect the results as this is another uncertainty source that is not taken into account?

If MDCA is better (at least taking into account that using together Tbh and Tbv adds 15 % of information) why SCA is the official SMAP algorithm and gives better results ? "There is strong interest in the MDCA approach because of its independent estimation of vegetation water status". I probably agree, but this if very very challenging using a single incidence angle. SMOS can do it because it provides multi-incidence angle Tbs. Konings et al. "How Many Parameters Can Be Maximally Estimated From a Set of Measurements?," in IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 5, pp. 1081-1085 have already explained that not because there are two measurements it is possible to actually retrieve two parameters.

Other comments ————

Line 18: raw data here is undefined. The authors should be more specific so that the abstract is self-explicative

Line 21: "inadequacy" is not a scientific term. What is that inadequacy? Where does it come from ?

Line 67: Peggy O'Neill et al. should be O'Neill et al.

Line 79: 0.04 m3/m3 accuracy target? Which is the metric the authors refer to ?

Line 136: The tau-omega model is not inverted at all. It is used as a forward model and the modeled Tbs are compared to the observed ones varying parameters such as SM. When the Tb's are similar to the observed ones, SM is assumed to be close the real value. There is no inversion of the model giving SM as a function of Tb.

Line 144: it is the uncertainty or the variable that is denoted as H(Yobs) ?

Line 150: The following sentence is meaningless "Although the detailed structure of best achievable model performance maybe remain unknown, mutual information, de-noted as I(XInputs; Yobs) where XInputs are the available inputs and Yobs is the in situ measured variable of interest, can provide a good benchmark measure". Please, rephrase.

Line 167: Eq. 2, what is the sense of writing an inequality comparing "mutual informa-tions" (I) with the uncertainty of the variable of interest (H(Yobs))? H and I should not be in the same inequality.

Furthermore, in the example of Eq.1 X is Tbs, Y is Ymodel and Z is Yobs as i) one measures the Tbs, ii) apply the model, iii) Compare to "ground truth". Therefore I(X,Y) >= I(X,Z) should be I(X_inputs, Ymodel) >= I(X_inputs, Yobs) instead of what is written in Eq. 2

Lines 175-180. The manuscript will be clearer if it is stated how to compute those quantities from the actual SM time series records (taking into account the uncertainties)

Line 193 Eq. 5 Why the "mutual information" is compared to uncertainties? Why uncertainties are assumed to be additive ?

Line 195: Eq. 5 expresses I as a function of HCN, how Hcn(Ymdca, Yobs) could be estimated by replacing anything in Eq. 5. Do the authors mean I(Ymdca, Yobs) can be ..." ?

Eq 10. What is -II ?

—- Typos Line 193 Eq. 5 TBv should be T_{B_v}.

Line 177. "is" and "the" are lacking. "Where p IS THE probability..."

Line 194: ... and H_CN() ARE the estimated joint ENTROPIES that ...

Line 196. It IS worth

Eq 8: U_1 should be U_2

Line 232: H_CN(h,v) should be H_CN(Tbh, Tbv)

Line 345. Please correct "theoretic"