

Interactive comment on “A new dataset of satellite observation-based global surface soil moisture covering 2003–2018” by Yongzhe Chen et al.

Anonymous Referee #3

Received and published: 24 July 2020

The authors used an iterative neural network approach to produce a new satellite based soil moisture dataset using 11 microwave soil moisture products, using SMAP data for training and ISMN database for validation. The approach is quite original and efficient resulting in an improvement in the accuracy of the spatio-temporal patterns at the global scale, and at a 0.1 degree resolution. However, the manuscript will need to be improved before acceptance, in its structure, clarity and tone.

1- The introduction would need to be improved. Several statements need to be supported by existing literature, others would need to be clarified. Finally the introduction would need to end with a brief description of the approach used in the study and how this approach will address the three major concerns raised from existing soil moisture products. See detailed comments below for details. 2- The tone of the manuscript

C1

when referring to the new product and to past studies is not always appropriate. For instance, stating that the present product is “superior to the existing products” is useless, not informative and condescending. I would encourage the authors to rather explain how their product is an improvement to the global estimation of soil moisture, without necessarily condemn other products. In the result section, while nonlinearities between estimate and in-situ soil moisture measurements are identified for other products, it is not reported for the author’s product which I find quite biased. 3- The validation approach is based on site specific comparison. However, soil moisture being so spatially variable within a 0.1 degree pixel, validation based on single site observations within 0.1 degree pixels can be quite meaningless. This might be particularly true when one considers that site selection for in-situ measurement is rarely motivated by representativity of the surrounding landscape, but by specific ecological reasons.

— Abstract. — “This new dataset, once validated against the International Soil Moisture Network (ISMN) records, is supposed to be superior to the existing products” → do you mean this validation hasn’t been done yet? Superior in what way?

“reveals that the surface moisture decline on rainless days is highest in summers over the low-latitudes but highest in winters over most mid-latitude areas.” Soil moisture being so spatially variable, I find the impact of this statement quite limited – e.g. low latitude regions range from tropical / equatorial rain forests to deserts and one would expect as much differences in the sensitivity of soil moisture to precipitation between a desert and a tropical forest than between a tropical forest and a temperate prairie.

— Introduction —

L47: “due to various disturbances”: what type of disturbances? L49: “ Although new sensors, SMOS . . .” -> “Although new sensors such as SMOS. . .” L 50: “better penetrability” -> please be more specific: what depth? L66: “Because the temporal variation in soil moisture is often better captured by model simulations than remote sensing inversions” : please include a reference that support this statement. L67: “CCI may

C2

undesirably combine the disadvantages of both.” Be more specific here (low accuracy of temporal variations from remote sensing products and low spatial accuracy from model simulations – am I right?). And please include another reference here for this second statement. L70: “are assimilated instead”: instead of what? this sentence is not clear. L85: “Among these three approaches, machine learning proves to be probably the best choice” based on what criteria – again, please be more specific. L102: “substantial success has not been achieved yet.” This is a rather strong and yet vague statement that denies the merits of a large body of research. Please remove this statement. L102/103: “the high-quality microwave observations are not fully utilized”: this is not clear from the literature review – please develop this point in earlier sections of the introduction (i.e. in what way high-quality microwave observations haven’t been fully utilized, and how the authors are proposing to utilize them more efficiently). L106-107: This statement should be removed from the introduction section.- this is rather a concluding statement. Instead please describe your approach in a couple sentences and how this approach addresses the three major concerns identified.

— Data and Methods — – Section 2.1 L110: please add citation to literature supporting this statement. L115-118: This sentence is too long and too complex. Please split into shorter and clearer sentences. – Section 2.2 L147/160: the purpose of this argumentation is quite unclear as the environmental predictors that are selected are also important drivers of soil moisture dynamic. L201: since precipitation is such an important driver of soil moisture, the reasons why this variable hasn’t been included as a quality impact factor should be included in the main document. L202/203: This sentence should start this section, not end it. A table similar to table 1 but for the quality impact factors would be useful. The table would indicate the source of the data, the resolution and the temporal span for the dynamic factors. – Section 2.3 Sections 2.4, 2.5 and 2.6 should be included in section 2.3 as it details the different steps of the calculation flow. A clear justification on why a neural network approach was adopted should be included in this section. L223: what is a hidden layer? L242: reference required for the “suspicious value removal”. L259: Here you are referring to a 1 degree

C3

pixel a presume ? Please specify. L268: why 74 networks? Please explain. – Section 2.7 Figure S2 should be included in the main manuscript as the spatial distribution of validation data is critically important to evaluate the overall strength of this validation. It is surprising that none of the Canadian sites or the Russian sites made it to the final set of sites for validation. L304/306: This is an important point. Soil moisture being so spatially variable, validation based on single site observations within 0.1 degree pixels can be quite meaningless. Especially when one considers that site selection for in-situ measurement is rarely motivated by representativity of the surrounding landscape, but by specific ecological reasons. – Section 2.8 L335: “probably the best choice for periodic function fitting” : please support this statement by adequate reference to literature. L344/345: I don’t understand this argument. Why restricting this analysis to 10 consecutive rainless days and not the whole range of 10-days sum of precipitation?

— Results — Tables 3 to 8 could be synthesized into only two tables: one for temporal accuracy assessment and one for spatial accuracy assessment for the three products comparisons with SIM. Similarly, it would be nice to have figures 3, 4, 6, 7, 9 and 10 summarized in 2 figures where all four products appear (SIM, SMAP, GLDAS and ASCAT). This would facilitate comparison between products.

L380-383: However, it looks like the relationship between SIM estimates and in situ observations is nonlinear (Figure 5a). Furthermore, SIM seems to overestimate soil moisture in the lowest range (winter?) when a density of pixels is quite high. Please include these remarks in the results.

L393: the relationship between SIM and in situ measurements is also obviously nonlinear. Please include this remark in the text for fairness.

Why table S19 and Figure S7 do not appear in the main document like the other product comparison? Please move them to the main manuscript.

L422-434: This belongs to the discussion section.

C4

