Author Comments

Response to Referee Comments on "Advances in Soil Moisture Retrieval from Multispectral Remote Sensing Using Unmanned Aircraft Systems and Machine Learning Techniques"

[hess-2020-271]

We are grateful for all the constructive comments and suggestions we received from both reviewers. In this document, we have copied the reviewer comments and placed our responses underneath each reviewer comment. Our responses are written in bold text and indented on both sides.

We sincerely thank the reviewers for their valuable comments.

Referee # 1 Comments and Author Responses

General Comments:

(1) The study presents results of soil moisture retrieval using unmanned aircraft system and four machine learning (ML) techniques. The authors conclude that BRT as a ML technique was better (3.8%) than RF (3.9%), ANN (4.3%), SVR (4.4%) and RVR (4.5%) even though the error bars (Figure 6) overlap. While I think the scope fits quite well within the scope of HESS, I think the authors need to improve the presentation and discussion of the results than is presented.

The differences in the methods in estimating soil moisture are infinitesimal to warrant such a claim considering that the error of TDR probes is 3%. The accuracy of the results would very much depend on the relationship of the training data set, presence of outliers as well as erroneous values in the training data.

> **We accept that the differences between the ML algorithms are small particularly considering the measurement accuracy. We will remedy this in the revised version by moderating our conclusion to indicate that all the ML algorithms yielded reasonable and comparative accuracies, but that RF and BRT algorithms have slightly higher accuracy. We will further highlight model simplicity in terms of training computational demand to set apart the algorithms (this favors RF) in the abstract and conclusion. This is already discussed in the results (Lines 440 – 445) but was not included in the abstract and conclusion sections.**

(2) In the abstract it is claimed that the UAS was used to create a high digital elevation model as well as quantify relative vegetation photosynthetic status. Yet photosynthetic 'status' results are not provided. It is also not clear what this means.

> **We agree that "photosynthetic status" is not an accurate term. In the revised version we will change this to "vegetation spectral indices".**

(3) TPI was observed to highly correlated with soil moisture. This would then mean that areas with more relief were well correlated with moisture. And by extension, NDVI (TTVI) would likely be well correlated with moisture in these areas. While vegetation indices were eliminated from list of variables used, how would vegetation index vary with moisture along the gradient of these convex and concave reliefs? (See: Ryan Engstrom, Allen Hope, Hyojung Kwon & Douglas Stow (2008) The Relationship Between Soil Moisture and NDVI Near Barrow, Alaska, Physical Geography, 29:1, 38-53, DOI: 10.2747/0272-3646.29.1.38)

> **This is an interesting question, and we would speculate that NDVI would likely be well correlated. It is likely that the ML learns the information contained in NDVI from the component bands. One strength of ML is that it learns underlying patterns whether they were known or not beforehand. Because NDVI (TTVI) was not used as input in our models we did not explore its relation with soil moisture.**

(4) The discussion of results misses out on comparing and contrasting the findings of the study with those existing in literature. The conclusion of the study is also very weak with no take home message.

Most detail is given on limitation of the study and future prospectus of the application rather than study findings.

> **We accept the need to expand the conclusion. In the revised manuscript we expand the conclusion with the main findings of the study including the advantages of the selected ML algorithms, the importance of terrain variables in addition to reflectance and hydrologic variables, and the relationship of important variables with soil moisture.**
>
> **The focus on future outlook is because this study was undertaken as a proof of concept. Although very promising, we wanted to highlight future directions that can be taken to improve the rigor and universal applicability of such models.**

(5) Figure 5 indicates that NDVI was at maximum at day 186 while moisture was low indicating previous moisture recharge was the likely cause of increased vegetation bloom. What was the role of lag time of vegetation indices (NDVI/TTVI) response to moisture recharge due to previous rainfall?

> **As you correctly identified lag time of precipitation is likely the main reason (on top of phenology). To account for possible lag time effects, we used rolling sums of antecedent precipitation at 1, 15, and 30-days (see Table 2 and Figure S) as inputs. Although we found these aggregated sums to be important predictors, our setup does not allow us to analyze deeper into the nature of the lag time relationship.**

(6) The BRT model prediction (Figure 8) shows under-estimation of high soil moisture content. The authors do not attempt to explain the reason for this underestimation especially for Julian day 186. Could this be because vegetation variables (NDVI/TTVI) were not included in the model?

> **You have correctly observed that our models underestimate high soil moisture and seem to cut-off around 40 - 45 % soil moisture content. Although NDVI might be the reason, we feel the more likely explanation is that there are few training data points with soil moisture above 45 % (Figure 5A). In the revised manuscript. We will add the following after line 458 to discuss this observation: "*The 1 to 1 comparison shows that our model prediction seems capped at around 40 % soil moisture content. This is most likely because there are very few training data with values above 40 % soil moisture (see Figure 5).*"**

(7) It is indicated in paragraph 235 that 12000 images were acquired. Can this be broken down to how many acquisitions per sample site were taken for each band and used for each of the 6 days UAS was flown?

> **We have edited the paragraph to remove the total number of images and instead explain the processing more clearly. The edited manuscript lines 237 – 242 now read: "*Images for the study area were captured with a minimum of 85 percent overlap and ground pixel resolution of 10 to 15 cm. Images were scaled to uniform 15 cm pixel resolution post-processing. Image processing was done using Pix4D photogrammetry software (Pix4D, Lausanne, Switzerland). Digital surface model (DSM) for the study area was photogrammetrically generated from the overlapping*"**

***stereo-images and images were orthorectified, and radiometrically calibrated. The DSM produced is synonymous with digital elevation model (DEM) for our study area given that no tall objects existed above the ground surface*"**

(8) Figure S9: There is no PET at 15-day aggregation.

That is correct, there is no PET-15-day aggregation, only 1 and 30-day aggregations (see table 2)

(9) Figure S5: Inference is made of NTWI being linearly correlated with log of acc (flow accumulation) whereas in the main manuscript text it is TWI.

NTWI is the normalized topographic wetness index so that the TWI values range between 0 and 1. It was normalized for aesthetic reasons to avoid large numbers in the figure and should not affect the correlation results. We will include the definition of NTWI in the caption for clarity.

(10) Use of TTVI index instead of NDVI is stated to eliminate negative values. Were there negative values of NDVI for the grassland dominated sample sites?

There were no negative NDVI values in our training data although there were some close-to-zero negative values in the study site raster. However, elimination of negative values was not the primary reason for using TTVI transformation of NDVI and we have decided to elaborate the explanation we gave on Line 292-293. In the revised manuscript we have added the following after Line 293: "… transforms NDVI histograms into a *more* normal distribution. *Such transformation of inputs is not required in the ML algorithms we used; however normalizing inputs is considered good practice and aids models to converge faster*."

Also relevant to the above comment, we will add an explanation at the end of Line 337 on the need to standardize as: "*Standardization is a common data preprocessing that eliminates the issue of scale among input variables and leads to better and faster training (Brownlee, 2018).*"

(11) Calibration panel was used to record stability of the UAS sensor for every flight for the six days and to compute surface reflectance. What were the uncertainty (errors) for the different days?

The calibration panel was used in the processing of radiometric correction to calculate absolute surface reflectance described in Section 3.3.1.2. The Pix4D software we used does not record calibration panel uncertainty in the processing report document and we are unable to get this information.

(12) Use of the word 'meteoric' and interchangeably with 'meteorological' variables when hydrological variable would be appropriate for the two variables used, precipitation and potential evapotranspiration.

We accept this recommendation and will use "hydrological variable" instead of meteoric or meteorological variables in the revised manuscript.

(13) The article has so many grammatical errors and would benefit greatly if the paper was re-written and organised accordingly to HESS standards. More comments are highlighted in the attached manuscript.

We will thoroughly revise the manuscript to address all grammatical errors and improve clarity. We will also address all the line-by-line comments.

Line-by-line comments

Line 73: revise this word. A better word than 'ill'-posed

We have decided to leave it as it is. "ill-posed" is commonly used in literature to describe problems that violate the requirements of a "well-posed" problem. We feel that other terms would not be as precise.

Line 80: Their

We accept the comment.

Line 88: More of what?

We have removed the "more" and re-written the last sentence as: "*A major drawback of analytical models is their complexity and their requirement for a large number of input parameters (Zhang and Zhou, 2016)*"

Line 148-149: Why?

There is no solid theoretical reason when it comes to deciding the number of hidden layers but most literature suggests a single hidden-layer is sufficient for most problems that are not extremely complex. Furthermore, multiple-hidden layers are very computationally demanding and difficult to train. We have re-written the last sentences to better communicate the reason why we chose the feed-forward neural network with a single hidden layer: "*In this study, we implemented the feed-forward neural network with a single hidden layer which is considered sufficient for the majority problems (Reed & Marks II, 1999)*."

Line 182: Acquired simultaneously during flights or separately? What was the resolution of the DEM?

The digital elevation model from the first day was used for this study. We have edited the sentence to clarify the resolution as follows: "*High-resolution, cm-scale, digital elevation model.*"

Line 199: State the size of area, rather than saying small subset

We have included the size of the area in parenthesis

Line 218: over

We have replaced "on" with "over" as suggested.

Line 218: Which dates? State the specific days.

We have now added the six dates in parenthesis.

Line 229: This is past mid-day. Better be clear and specific because in the beginning you state it is mid-day

We have now modified line 227 to read: "… late mornings to early afternoon".

Line 237: Can this be broken down to how many acquisitions per sample site were taken and used for each of the 6 days UAS was flown?

We have edited lines 237-242 to clarify these methods. The edited manuscript lines 237 – 242 now read: "*Images for the study area were captured with a minimum of 85 percent overlap and ground pixel resolution of 10 to 15 cm. Images were scaled to uniform 15 cm pixel resolution post-processing. Image processing was done using Pix4D photogrammetry software (Pix4D, Lausanne, Switzerland). Digital surface model (DSM) for the study area was photogrammetrically generated from the overlapping stereo-images and images were orthorectified, and radiometrically calibrated. The DSM produced is synonymous with digital elevation model (DEM) for our study area given that no tall objects existed above the ground surface*"

Line 238 Did the resolution vary this much? Was it not standard and fixed?

The resolution of the orthorectified images varied slightly between the different dates. However, pixel resolution was standardized post-processing to 15 cm.

We have re-written this paragraph. Please see our response to the previous comment.

Line 238: Did the sensor also collect stereo-pair images for DEM development?

Yes, this has now been explained on line 182. Please see the response to the comment on line 182.

Line 249 – 250: State the reason for doing this?

We have added the following in the revised manuscript: "Doing so ensured that our samples included a variety of landforms."

Line 255: Was this conducted simultaneously during the flights? Or was residencely installed and collected data continuously?

This was done simultaneously during flights. Please find a more detailed description on line 242.

Line 292 – 293: Wondering why you would expect -ve NDVI values over the flight grassland terrain?

There were no negative NDVI values in our training data although there were close to zero negative values in the study site raster. However, we have decided to elaborate on our choice of TTVI since the explanation we gave on Line 292-293 is not clear. In the revised manuscript we have added the following after Line 293: "… transforms NDVI histograms into a *more* normal distribution. *Such transformation of inputs is not required in the ML algorithms we used; however normalizing inputs is considered good practice and aids models to converge faster*."

Also relevant to the above comment, we will add an explanation at the end of Line 337 on the need to standardize as: "*Standardization is a common data preprocessing that eliminates the issue of scale among input variables and leads to better and faster training (Brownlee, 2018).*"

Line 306: hydrological

We accept this recommendation and will use "hydrological variable" instead of meteoric or meteorological variables in the revised manuscript.

Line 339: This is strange. TWI is well known to correlate well with soil moisture.

It is likely that the role of TWI is scale dependent. While TWI is important controlling variable at larger (e.g. watershed) scales, it is likely less so at this smaller scale area.

Line 344: What does this mean?

We will address this by adding a parenthesis that will refer the reader to Section 3.3.2.2 for more details on the calculation of TPI.

Table 3: Julian day? Were these the same days in which UAV was flown also?

These are days into the water year (i.e., starting at 1 for October 1st). Yes, the flight and sampling dates are the same.

Table 3: Are these 1-day, 30-day cumulative sums or annual? What do you mean by water-year?

These are cumulative sums of antecedent precipitation starting from the start of the water year (i.e., October 1). By water-year, we mean the hydrological year. For California, water-year starts on October 1 and ends on September 30 of the next year.

Line 388: This is 27.8%

Correct. The split was not exactly 25 % since it was done by transects, not by the individual points.

Lines 429 – 433: Repeated. Has already been described in text in methods description sections

We accept this comment. We have now removed the entire section.

Line 436 – 437: Training or validation results?

We have now edited it to include that the accuracy was in the testing set.

Line 449: ??

We have corrected the grammar error as: "… were selected at random for each…"

Figure 6: What is XBRT?

This is supposed to read BRT. We have changed the inconsistent abbreviation to "BRT".

Figure 8: Prediction underestimates high soil moisture.

We agree with the observation that our model underestimates high soil moisture and seems to cut-off around 40 - 45 % soil moisture content. Although NDVI might be the reason, we feel the more likely explanation is that there are few training data points with soil moisture above 45 % (Figure 5A). In the revised manuscript. We will add the following after line 458 to discuss this observation: "*The 1 to 1 comparison shows that our model prediction seems capped at around 40 % soil moisture content. This is consistent with the lack of training data with values above this (see Figure 5).*"

Figure 9: Describe the initials in the y-axis in the caption. What is Cur?? Is ET equivalent to PET?

We have now described all the abbreviations in the caption including Cru which stands for curvature. We have also fixed ET to be PET.

Line 481: Contradicting. Decrease as surface becomes less convex (meaning tending to concave). What is the explanation for this?

By "less convex" we mean as the surface degree of convexity decreases. That is the surface is still convex but less so, it is closer to a flat surface. This is counter-intuitive. We have added the following possible explanation in discussing this in the revised manuscript: "*a possible explanation for this could be that more flat surfaces (with near 0 curvature value) are associated with higher slope areas, as in the sides of slopes, compared to surfaces high in convexity which might be located near ridgetops which have relatively shallower slopes.*"

Line 489: ??

We have fixed the language error.

Line 489: This would then mean that areas with more relief were well correlated with moisture. And by extension, NDVI (TTVI) would likely be likely well correlated

Very likely. However, NDVI was found to be not important in the models likely due to the inclusion of the constituent red and near-infrared bands.

Line 490: Precipitation and Red band had the highest importance

Correct. We have now clarified the statement as: "*Of the TPI values calculated at different scales, the scale that had the highest importance for soil moisture prediction was TPI with 15 and 35 m inner and outer diameters, respectively (TPI(15,35))*"

Line 500: This is because of lag time of vegetation response following moisture replenishment through rainfall

Yes, we believe this to be a major driver. One of the reasons we used running sums of rainfall on different days was to capture this aggregated effect of preceding day rainfall.

Line 500: prediction

We have corrected the typo.

Figure 10: The plots could be labelled a), b) etc. and while discussing in text reference should be made accordingly

We accept this comment and have labeled them alphabetically in the revised manuscript.

Section 5 Conclusion and outlook: What was the conclusion from the study results and findings?

We accept the need to expand the conclusion. In the revised manuscript we expand the conclusion with the main findings of the study including the advantages of the selected ML algorithms, the importance of terrain variables in addition to reflectance and hydrologic variables, and the relationship of important variables with soil moisture.

The focus on future outlook is because this study was undertaken as a proof of concept. Although very promising, we wanted to highlight future directions that can be taken to improve the rigor and universal applicability of such models.

Line 530: But the grasses have shallower root zones

Yes, this is true. In the context of the study area, we consider a shallower root zone.

Referee # 2 Comments and Author Responses

This research provides an excellent example of soil moisture retrieval from multispectral UAS remote sensing using machine learning methods. The manuscript is well written and fluent for the reader in general. Introduction of the manuscript is well-designed with convincing literature review and clear objectives; The description of dataset and methods is concise. The presentation of results is well organized with logical steps. The conclusion part is well organized with valuable information and suggestions.

> **Thank you for the positive comment.**

General issues:

First, is there any possibility to validate the final soil moisture map excluding the sampling points used in the model training? Although the predicted soil moisture map looks reasonable, the accuracy is not guaranteed without a validation.

> **Yes, the final soil moisture map can be said to be entirely made of data not used in the training. The training data are only points compared to the extent of the map.**

Second, sometimes the figure name in the text is "Figure 1", sometimes it is "Figure S1", please check throughout the text.

> **The figures that start with "S" are figures in the supplemental document. To avoid confusion, we will note this in the first instance of this numbering format in the revised document.**

Minor issues:

Line 126, ". . . low operating costs (Anderson and Gaston, 2013; Berni et al., 2009; Colomina and Molina, 2014; Elarab, 2016)". You can also add some new references, such as: (1) Manfreda, S.; McCabe, M.F.; Miller, P.E.; Lucas, R.; Madrigal, V.P.; Mallinis, G.; Dor, E. Ben; Helman, D.; Estes, L.; Ciraolo, G.; et al. On the use of unmanned aerial systems for environmental monitoring. Remote Sens. 2018, 10, 641. (2) Tmuši´c, G.; Manfreda, S.; Aasen, H.; James, M.R.; Gonçalves, G.; Ben-Dor, E.; Brook, A.; Polinova, M.; Arranz, J.J.; Mészáros, J.; et al. Current Practices in UAS-based Environmental Monitoring. Remote Sens. 2020, 12, 1001.

> **Thank you for the suggested citations. We will add these references in the revised manuscript.**

Line 146, the abbreviation "PTFs" appeared without definition. Please check throughout the text.

> **We accept this comment. We will expand the abbreviation in its first instance.**

Line 293, The reason why "transforms NDVI histograms into a normal distribution" is an advantage in this research may need to be explained here.

> **We will add the following explanations as suggested after Line 293: "transforms NDVI histograms into a *more* normal distribution. *Such transformation of an input is not required in the ML algorithms we used; however normalizing inputs is considered good practice and may aid models to converge faster*."**

Also relevant to the above comment, we will add an explanation at the end of Line 337 on the need to standardize as: "*Standardization is a common data preprocessing that eliminates the issue of scale among input variables and leads to better and faster training (Brownlee, 2018).*"

Line 297, why the original resolution of DEM is 6.85 cm, while in line 238, it mentioned that the pixel resolution of the captured images are 10 to 15 cm"? As the DEM map should be generated from the same stereo-images, they may share the same resolution.

The DEM was generated from the fifth, high-resolution visible channel which had higher resolution than the multispectral channels. We will edit the sentence to remove this apparent discrepancy as: "*For this, we first upscaled the DEM to 15, 30 …*". We will further edit lines 237-242 to clarify the DEM production. (See response to Reviewer #1 general comments number (7)).

Line 323, an explanation on the reason why a "standardization" is needed here and how does it benefit on the model training could be added. Seems this procedure may eliminate the physical meaning of variables.

We will add an explanation at the end of Line 337: "*Standardization is a common data preprocessing that eliminates the issue of scale among input variables and leads to better and faster training (Brownlee, 2018).*"

Line 357, "April 4, 2018", replace the label "Day of the Water Year" of x axis in Figure 5 with exact dates could be better.

We feel that since day of water year is used on all subsequent relevant figures it should remain the same however, we have added the day of year for April 4, 2018 in Line 357 for clarity.

Line 361, ". . . some terrain variables. . ." could be replace with "vegetation index", which is the only variable presented in Figure 5.

Figure S5 actually refers to a figure in the supplement that has other terrain variables. To avoid confusion, will note that the "S" refers to supplemental material in the first instance of this numbering format in the revised document.

Line 365, ". . . variables selected variables in the data is shown in Figure S6": duplicated word "variables"; the content is not in

Similar to the comment above, this S6 is a figure in the supplemental material. In this instance, we will further clarify by mentioning this is supplemental material as: "… shown in Figure S6 *in the supplemental material*."

Figure 6. Line 451, "XBRT" should be "BRT".

We accept this comment. We will replace all instances of XBRT with BRT.

Line 459, a detailed explanation of Figure 8 is necessary.

We have re-written the description as follows: "*A Scatter of the measured versus predicted soil water content of the testing sets around 1:1 line. MAE, MBE, and R2 are averaged across the 30 models.*"