Specific comments

Line 22: Development of simple accessible metrics to assess soil health facilitate spatial and temporal sampling density but should also support the implication of farmers, consultants and even citizen in soil health assessment.

Sentence has been reworded to incorporate suggestions.

Line 28-29: The degree of slaking determines if the process produces a favourable or unfavourable environment for cultivation and plant growth. It is true but not sufficient. It also determines the degree of soil conservation because the aim is to cultivate but likewise to protect this resource.

Added "and has implications for soil conservation."

Line 45: The authors focus on agricultural practices that increase soil susceptibility to slaking, but what about practices limiting susceptibility to slaking? Carbon management, crop successions, superficial or "light" tillage. . .

Added "Techniques that increase soil organic matter such as cover-cropping, reduced tillage and application of organic amendments may reduce susceptibility to slaking"

Line 59: Another group of widely used method to estimate aggregate stability (that is the contrary of slaking) is the Mean Weight Diameter (MWD) after wet sieving of soil aggregates. You should mention this reference method.

Added "Established methods to quantify stability of aggregates subject to wet-sieving (Yoder, 1936) or simulated rainfall (Schindelbeck et al., 2016) are also time-consuming and require specialist equipment."

Line 93-95: Please, provide the equivalent of soil references according to the World Base Reference for soil classification.

Australian Soil Classification has been removed and text changed to: The soils of the floodplain area at L'lara are classified as Vertisols according to the World Reference Base for Soil Resources, with some expression of calcic horizons (IUSS Working Group WRB, 2015). The sand hill area is represented by Luvisol, Lixisol, Solonetz, Leptosol and Regosol soil groups.

Line 96: Please define "broadacre".

The term 'broadacre' has been removed to prevent any ambiguity as the term has limited use outside of Australia. The remaining sentence still conveys the same meaning.

Line 96: Is L'Iara covered with a soil map? If yes it and if it is relevant, it could interesting to add this map (near figure 1 for instance). It not, a land use map could also be helpful to interpret figures 6 and 7.

A soil type map was produced but unfortunately only covers L'Iara and it is in the Australian Soil Classification and we have been requested to use WRB. MrVBF, NDVI and land use maps have been added to Figure 1. The MrVBF map gives a good indication of the distribution of Vertisols versus other soil with a sandy topsoil.

Line 111: "in the area surrounding L'Iara, an additional 50 samples. . ." or "50 additional samples"?

Text has been modified at request of RC2.

Line 108-119: Sampling scheme: collection of datasets with various sampling approach. I guess they came from various field campaigns and programmes. What are the dates for each one? A summary of the distribution of land use at the observation points is missing. It could be a table or a sentence in the text.

Text has been modified for clarity and a table added summarising the sampling dates and number of observations for each land use for each campaign.

Line 121: What was the size of the 20 to 30 aggregates? I suppose that it was for each soil sample. Please mention that.

Target diameter of "(ø 5-10 mm)" given in text.

Line 130-132: These 2 sentence could be move to the 2.3 section and replace the 2 first sentences of this section. I suggest renaming this 2.3 section: "soil sample preparation and laboratory methods" (or something like that).

Lines 130-132 moved to section 2.3 and section 2.3 renamed "Sample preparation and laboratory methods". Note number of samples changed to "12 to 15" at the request of RC2.

Line 141: '10 minutes'

Existing grammar is correct, suggestion not incorporated.

Line 145: It the difference between replicates was more than one, only the unique additional reading was considered for the final result of SI? And what would happen if this additional reading was an outlier one? How many times a third observation was necessary?

Modified text: "An additional reading was required for approximately 20\% of samples and was more commonly required for soils with higher slaking index values compared to samples which exhibited minimal slaking. When additional readings were taken the outlier reading was discarded and remaining readings averaged to provide the final SI for each sample."

When an additional reading was taken it was always within one unit of one of the original duplicates. The additional sample and the duplicate within one unit were then averaged to give the final slaking index value and the other duplicate was treated as an outlier and not used in the calculation.

Line 160: Please name other approaches.

Text has been modified "unlike the ASWAT test that requires 2 hours of immersion."

Line 174: All terrain attributes are not at the same spatial resolution. Slope, aspect, MrVBF and MrRTF could have been obtained from the 5m DEM since it was available.

The 5 m photogrammetry DEM provides the most accurate point estimate of elevation but it is not hydrologically enforced and for this reason we prefer to use the elevation derivatives calculated from the 30 m SRTM DEM.

Added "and gives an accurate point estimate of elevation though it is not hydrologically enforced".

Line 178: Why potassium concentration is of particular interest?

Added "Variation in the concentrations of the radioelements are indicative of change in soil type or parent material".

Line 184: How was made the split between training and test datasets?

Text has been updated in section 2.2. to clarify this.

Line 190: "The kriged residuals was were added. . .". There is non information in the text about the variogram of the residuals? Were residuals spatially structured?

Added information about kriging of the residuals: "The residuals (difference between the observed and predicted SI values) at observation points showed a weak spatial autocorrelation. A Gaussian

Line 196: The first sentence is not clear. Please reword. You could also rephrase the second sentence. Line 198: Observation points are allocated into classes having similar behaviour. How many classes? How the choice of classes and allocations of observations was done?

Start of paragraph reworded "Relationships between SI and measured soil properties were explored to identify potential contributing factors as a means to inform management practices to reduce excessive slaking. Two classes of soils were evident in the samples, soils with clay content ≥25% and CEC:clay ratio≥0.5 which consistently exhibited excessive slaking, and other soils.

Table 2: It would be relevant to distinguish training and test datasets to confirm that they cover a similar range of soil attributes values, especially because of the difference in location between the 2 datasets: training data only located within L'Iara boundaries.

As indicated in Figure 1 the test set is located entirely within L'Iara and is inter-mixed with samples from the training set. In this instance I don't believe it is necessary to confirm that the samples occupy the same covariate space.

Line 225: ". . .in these samples. . ." which ones? With clay content >25%?

Text modified "The majority of clay soils had a high CEC:clay ratio indicating that the dominant phyllosilicate in the clay soils studied is smectite."

Figure 2: It would be useful to know the number of samples in each of the classes land use/clay by adding this information in the figure. What about statistical significance of the differences between classes?

The number of observations for each class and significant differences (p<0.05) between means calculated using Tukey's HSD have been added to the plot and discussed in the text.

Line 267: I guess "3)" has to be suppressed.

"3)" was an incomplete reference to "(Fig. 3)". Corrected in text.

Line 301: The scenario of an increase of SOC by 1% conduces to predict a reduction of SI of 1.59 units for soils with clay content >25% and CEC:clay ration >0.5 according to the decay function. Values of SI depending on OC are widely dispersed around the model (figure 4). Nevertheless, the map of

change in SI after increase of C is based on this weak model. I suggest the authors to be more cautious in their conclusions concerning the effect of OC change on SI. Some elements of discussion about uncertainty are expected.

Added – "provided a weak fit to the available data"

Additional discussion points added to section 3.2.4 - "Another contributing factor for the improved validation metrics under increased OC scenarios is due to the SI values being based on modelled data which has had all unexplained error removed. Future efforts should account for the error of the underlying regression equations and quantify the uncertainty of the resultant maps by bootstrapping and applying random error based on the the prediction variance of the underlying regression equations."

Line 321: ". . .for some models". How many models were run? Please complete the section 2.6.

Changed to "in the model". A single cubist model was calibrated and LOOCV used for validation.

Line 345: 'patterns'

Changed to 'features'

Line 353-354: The accuracy of the mapping process was assessed, but not the real effect of increasing SOC content by 1% because uncertainty of the decay function of SI with SOC (the map was based on) was not estimated. This must be specified to avoid misunderstanding of this result.

Stipulated that the validation metrics refer to the "mapping procedure" and also added "Another contributing factor for the improved validation metrics under increased OC scenarios is due to the SI values being based on modelled data from which unexplained error has been removed."

The authors would like to thank RC1 for their constructive review, we have also simulated the change in slaking index under a 0.5% increase in OC as suggested.