AUTHOR RESPONSES TO REVIEW #3

REVIEWER #3: 1. Quantity and quality of the measurements: Spatial representativeness and significance of the statistical analysis

AUTHOR RESPONSE: In the discussion of individual points below we address these topics:

Quantity of the measurements: Spatial representativeness

- The key part of the pilot was the testing of the sensors in a harsh tropical environment, compared to other temperate environments that the authors had previously deployed sensors within. The rigorous data cleaning developed, while it reduced the number of data points available for model development, did show
  - the types of issues with such sensor deployments,
  - how they might be overcome in post-data collection filtering, and
  - how they might be avoided during data collection

Quality of the measurements: Spatial representativeness

- The goal of our pilot deployment was not to show that we could capture the heterogeneity of tropical pastures with just 2 sensors, but to show that it was viable to use such sensors for pasture monitoring. In the paper we discuss how ideally there would be a number of sensors deployed which capture the pasture heterogeneity
- The field sampling was designed using standard techniques that are used in this environment, and using research staff. As we discuss in the points below, the biomass measurements were made using photo standards that are the industry standard for visual assessment of pasture status in this environment. In temperate pastures the authors have used a spatially hierarchical method of pasture cuts to measure pasture biomass, but in northern tropical pastures this would not be appropriate.
- We also discuss below why sampling was not made in the area around the sensor footprint, which was by design as the spatial heterogeneity of these pastures (see example photo below) means that the surrounding measurements are not likely to be representative of the small FOV of the sensor.
- This spatial heterogeneity of the pastures raises the question of the sensor being able to represent this larger area. In the discussion the authors address this, by reinforcing that the FOV of the sensor is what is being calibrated in this pilot, and that to sample a full paddock area it would be necessary to deploy a number of sensors, likely of lower quality/cost, arranged spatially so as to capture the particular spatial pattern of a future location.

Significance of the statistical analysis

- We illustrate in the paper that we would not be using the models we developed for later prediction, but only to assess the strength of relationships as part of the pilot project. If we could show that after all of the data quality control that it would still be possible to develop relationships between the sensor and field data, then this would be the basis for further work to extend the relationships.

REVIEWER #3: 2. Statistical issues not entirely addressed as is the case of collinearity analysis and independent validation.

AUTHOR RESPONSE: In the discussion of individual points below we address these topics:

For brevity some of these topics were only briefly addressed in the manuscript. We have provided some of this extra detail (below) and at the editors discretion we can expand on these topics in the revised manuscript.

Collinearity analysis:

- The authors did address linearity in the exploratory data analysis of the large field dataset prior to statistical analysis. It was to avoid this problem of collinearity that the authors controlled the number and type of variables in the model development by:
    - Never using two sensor variables in a model.
    - Never combine two climate/environment variables such as soil moisture and precipitation.
    - For the 2-variable models, combining only one sensor variable with one climate/environment variable.
    - Not including some variables as they were found to be too highly correlated with other variables and so would not add to the analysis.
- As part of addressing the reviewer's comment we re-examined our initial exploratory data analysis, including the scatterplots and correlation matrix (see below) and found that the spectral variables can be considered to be too highly correlated with four of the climate/environment variables. Only the minimum temperature and the rainfall should be included together with the sensor data in the models. This is an oversight in our data analysis and we thank the reviewer for suggesting that we re-visit these earlier analyses. In the revised manuscript we therefore will remove this subset of 2-variable models that show this collinearity, and change the results discussion, tables, etc. to reflect this.
- The the key one-variable models are not impacted by collinearity or this issue. This change will not impact on the most important relationships discussed in the paper, which are the single variable relationships between the sensor and field measurements.

Independent validation:

- **We** agree with the reviewer that a bootstrapping or other similar approach would be an option for performing some validation. However, given our limited dataset we did not implement such a method as we were concerned about the sample size of the dataset once we excluded all but the wet-season data. Of the 112 field samples, and the 63 times where we had matching field data and at sensor data from either the multispectral sensors or the digital cameras, only 23 remain when we look at just data from the wet season. Of these, only 10 matching data pairs remain where there is a matching multispectral measurement. We considered a sample of 10 to be too small for bootstrapping or the "leave one out" category of methods.

REVIEWER #3: 3. Indirect estimations of ground truth where potential errors have not been evaluated. I found very interesting (and useful for potential readers) the discussion on the technical limitations of the system and the recommendations and "lessons learnt" that can be derived. However, I am quite skeptical about both the methods and results described in the paper regarding the relationship between sensor data and field estimates. Having said this, I would recommend refocusing the manuscript in order to put more emphasis in those issues related with sensor deployment, technical problems, data filtering, spatial and temporal trends found, etc. and reduce, or even remove, the model development unless more robust relationships can be established. Specific comments addressing particular scientific/technical/formal issues follow:

AUTHOR RESPONSE: We have attempted to detail the limitations of having only two nodes, sufficient matching data for model development, and capturing only a limited amount of the variability within the paddock. Particularly as the 63 pairs of matching field-sensor data were reduced to only 10 pairs during the wet season after the initial analysis showed that outside of the wet season the relationships were unlikely to hold.

The intent of developing these relationships is to show that future deployments of sensors _are_ possible even with our limited data, and that despite the necessary data filtering we still managed to achieve encouraging results. If we were to exclude the modelling analysis then it would also exclude the value of showing that developing such relationships are possible, and also suggesting where the effort of future work should be focussed (i.e. during the wet season). These key points could be further emphasised in the abstract and conclusion.

REVIEWER #3: Abstract. The last part of the abstract, where the information on the model results is provided, is not easy to understand if you haventt read the whole paper. I would recommend to rewrite this part to simplify the information and provide only the main results without explicit references to dates and variables that have not been previously explained and, therefore, are difficult to contextualize.
Introduction

AUTHOR RESPONSE: This section can be addressed in a revised version of the manuscript.

REVIEWER #3: In this section I miss references to specific studies using proximal-remote sensing to monitor grassland ecosystems and more specifically those related with on-farm decision-making analyzing the state of the art regarding methods and instruments so the reader can easily get the main scientific contributions of the paper.

AUTHOR RESPONSE: In the revised version we suggest adding some more specific references as suggested. For example:

> "Previous research using proximal sensing of pastures aimed at helping decision making in livestock production has employed handheld active multispectral sensors to measure green herbage mass and predict pasture growth rate (Trotter et al. 2010), measuring plant height (Payero et al. 2004), hyperspectral handheld device for nutrient composition (Pullanagari et al 2012) and monitoring pasture variability using multiple sensors (Serrano et al., 2016), Forage Biomass (Flynn, et al, 2008) and Quality (Zhao et al, 2007). These proximal sensing devices can certainly aid in farm management such as grazing and livestock nutritional management however they are time consuming for the producer which reduces the temporal frequency with which they are used to collect data for timely decision making. Proximal sensors deployed permanently in pastures can provide frequent information of temporal changes for timely management. These sensors may prove useful in livestock production under grazing conditions when decisions have to be made frequently (cell or rotational grazing) or at critical decision making periods such as during transitions of seasons."

- Trotter, M., Lamb, D., Donald, G., & Schneider, D. (2010). Evaluating an active optical sensor for quantifying and mapping green herbage mass and growth in a perennial grass pasture. Crop and Pasture Science, 61 (5), 389 -398.
- PAYERO, J.O., NEALE, C.M.U. and WRIGHT, J.L., 2004, Comparison of eleven vegetation indices for estimating plant height of alfalfa and grass. Applied Engineering in Agriculture, 20, pp. 385–393
- Pullanagari, R.; Yule, I.; Tuohy, M.; Hedley, M.; Dynes, R.; King, W. 2012. In-field hyperspectral proximal sensing for estimating quality parameters of mixed pasture. Precision Agriculture 13: 351-369.
- Serrano, J. M., Shahidian, S., & Marques da Silva, J. R. (2016). Monitoring pasture variability: optical OptRx® crop sensor versus Grassmaster II capacitance probe. Environmental Monitoring and Assessment, 188(2), 1-17. doi:10.1007/s10661-016-5126-5
- Zhao, Duli, Patrick J. Starks, Michael A. Brown, et al. 'Assessment of Forage Biomass and Quality Parameters of Bermudagrass using Proximal Sensing of Pasture Canopy Reflectance', Grassland Science, vol. 53 (1), (2007), pp. 39-49.
- Flynn, E. Scott, Charles T. Dougherty, and Ole Wendroth. 'Assessment of Pasture Biomass with the Normalized Difference Vegetation Index from Active Ground-Based Sensors', Agronomy Journal, vol. 100/no. 1, (2008), pp. 114.

REVIEWER #3 In page 18009 lines 15-16. I recommend refining this phrase as, in some cases, the operational availability of satellite derived products for decision making is not "so expensive". That basically depends on the application including the combination of spatial and temporal requirements.

AUTHOR RESPONSE: We are specific in this sentence to refer to operational delivery of pasture status from remote sensing, based on the work in the reference article. While satellite data is widely available at low cost for the coarse resolution products, high resolution images remain costly, and there are extensive operational costs in running a system which take raw satellite images, convert them to a variable such as a pasture growth rate, and deliver results to a farmer in a tailored way that can be used for decision making on farm. What could be added at the end of this sentence would be to refer to the availability of cheaper data that might not be fit for use:

> "Converting raw satellite images to a measure that is useful for on-farm decision making is also problematic due to the cost and processing requirements for operational delivery (e.g. Handcock et al., 2008), **and while cheap or free satellite images are increasingly available, their ability to be interpreted for decision-making on farm is not straight forward**."

REVIEWER #3: In page 18012 lines 15-16. It would be very helpful to add here information on the min and max FOV obtained (according to vegetation height changes).

AUTHOR RESPONSE: While the full FOV is not a critical parameter since the area of interest in the scene was always captured by the camera, we could add the following:

> "The camera field of view was approx. 2.8 m x 2.0 m at ground level, and would have been able to capture the 1 x 1 m area with a vegetation height up to approx. 1.5 m."

REVIEWER #3: PAGE 18012 Methods. Same page lines 19-24. Authors should explain here why they took the decision of having one fenced and one unfenced node (implications in methods and results). This is not clear neither in the methods or the results/discussion sections

AUTHOR RESPONSE: The decision was made to place the nodes in grazed and ungrazed paddocks so that the vegetation that was observed in each was at different heights. Although the grazed paddocks did have beef cattle in the paddocks for short periods during the sensor deployment, due to the lack of feed in the paddocks there was no discernible difference in vegetation height before and after the grazing. While we have left in the reference to grazing in the title and text to make it clear that the pastures in this study are used for beef production, this could be confusing.

To clarify these points we suggest that in the revised manuscript we add this point to page 18012, line 19-24, and changing the reference to the "grazed pastures" in the title and page 18011 line 16 to be "grazing pastures". This is to make clear that the study is focussing on grazing pastures that are used in beef production, and not on the difference between grazed and ungrazed pastures.

REVIEWER #3: PAGE 18013 Methods. In page 1803 section 2.2. Please check the use of the acronyms. I suppose here the correct acronym should be always VWC and not VMC?

AUTHOR RESPONSE: This is a typo and we will correct this in the revised manuscript. The Decagon sensor documentation uses VWC.

REVIEWER #3: PAGE 18013 Methods. Same page line 25. Please, indicate the correlation found

AUTHOR RESPONSE: When the available weather stations were compared with the soil moisture data, the average of "Charters Towers Airport" and "Townsville Airport" stations had the strongest Pearson product-moment correlation coefficient of 0.61 during the wet season period of data collection, and 0.40 for all data in the wet and dry season.

REVIEWER #3: In page 18014 lines 13-14. Same FOV? Do you mean that the two instruments where somehow aligned? Same size and area observed?

AUTHOR RESPONSE: The cameras were centred over the area sensed by the Skye sensors but does not have the same FOV. We could clarify this in the text as by adding to this sentence as follows:

"At each of the two nodes we deployed a Pentax Optio WG-1 digital camera in a downward-pointing position **centred on the area sensed by the Skye sensors so that images covered** the same FOV as the multi-spectral sensors."

REVIEWER #3: PAGE 18014 Methods. Same page line 18. Please, explain why automatic balance was chosen while most authors and networks (as is the case for Phenocam http://phenocam.sr.unh.edu/webcam/) recommend manual/fixed white balance in order to minimize in-camera processing of day to day changes in illumination.

AUTHOR RESPONSE: Please see the response on automatic white balance in our response to the previous manuscript comments, as well as our decision to use images from the middle of the day in order to minimise illumination changes between images.

REVIEWER #3: In page 18015 line 8. Please, explain why the image taken nearest to 12:00 was selected. In general all these decision should be better justified.

AUTHOR RESPONSE: For reasons of manuscript length we have condensed the detail on the deployment methods. The image nearest 12:00 was selected to minimise shadows and to ensure as consistent illumination as possible.

REVIEWER #3: PAGE 18015 Methods. Same page line 13. Manually here means visually?

AUTHOR RESPONSE: Points were visually assessed individually using the "Calibrate threshold" function in the VegMeasure software.

REVIEWER #3: PAGE 18015 Same page lines 16-17. Is this crosscheck analysis somehow included in the results?

AUTHOR RESPONSE: See results in Table 5 where the GLA was compared to various fractions.

REVIEWER #3: PAGE 18015 Same page line20. Was the FOV verified/characterized by the authors before sensor deployment. If not, it would be convenient to explicit that FOV is 25 degrees "as indicated by the manufacturer" considering that published works have demonstrated that some sensor FOVs greatly differ from manufacturers specifications. See the work of MacArthur, A.; MacLellan, C.J.; Malthus, T., "The Fields of View and Directional Response Functions of Two Field Spectroradiometers," in Geoscience and Remote Sensing, IEEE Transactions on , vol.50, no.10, pp.3892-3907, Oct. 2012 In general, all information about instrument calibration/characterization before its deployment in the field is quite useful and should be explained.

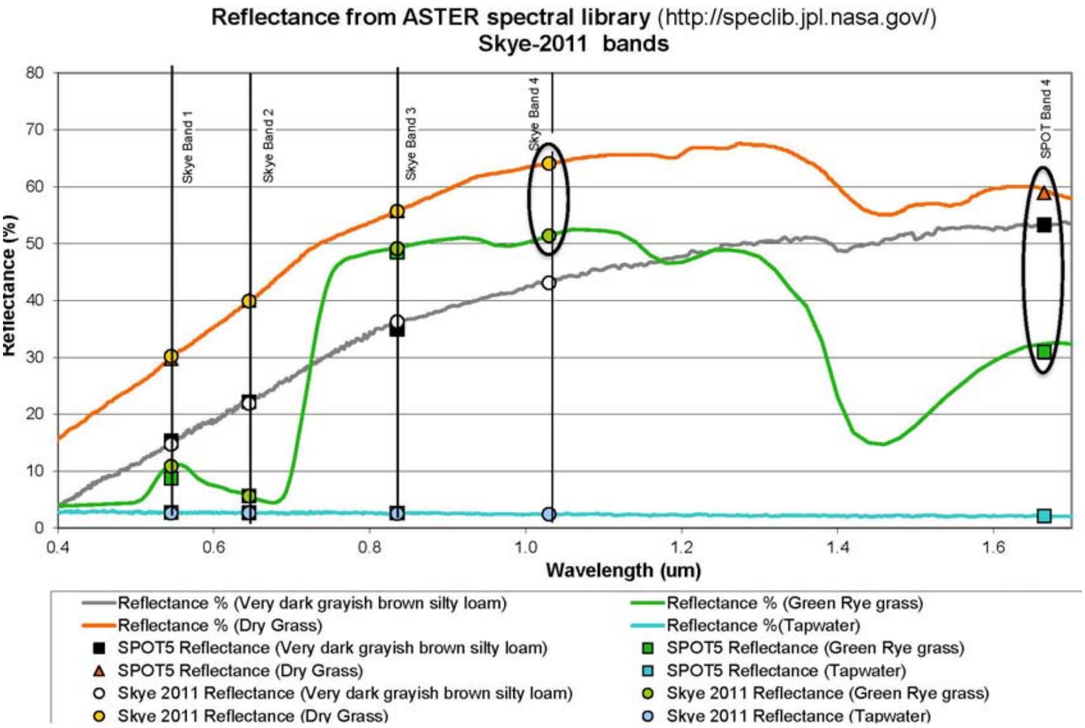
AUTHOR RESPONSE: We did not test the FOV of the sensors before deployment given that the sensors were calibrated individually by Skye and created to our specifications. This should be clarified in the text however, to add the suggested "as indicated by the manufacturer".

REVIEWER #3: Methods. In page 18016 lines 5-8. I recommend refining this phrase as the SWIR bands sensitive to vegetation water content are not the ones selected in this study.

AUTHOR RESPONSE: The reviewer is correct in that our final band selection was not in the *1.55–1.75 μm* spectral region recommended by Tucker., but instead in the longest wavelength that was possible to capture, as best as possible, senescing vegetation. For brevity some of this detail (described below) was lost, and we agree with the reviewer that it would be beneficial to expand this description here to make it clear the decisions behind the final band selection. This would also link better to the discussion section (page 18025 line 18) where we currently say "

"The lower SWIR band of our multi-spectral sensors was also in the lower part of the SWIR range, which is not as responsive to dry vegetation as the longer SWIR bands."

We originally planned to have the fourth sensor band located in the SWIR region to pick up dry vegetation, based on previous work in temperate dairy pastures, but due to design limitations of that model sensor it was not possible to have that long a wavelength. This resulted in the band choice in the lower SWIR (1.028 to 1.029 µm) spectral range. When we selected the bands, we examined whether this 4th band was likely to be suitable for monitoring dry pastures by comparing the band choice to the ASTER spectral library. The figure below shows the results of this analysis. The ellipses highlight the difference between the spectra of the dry and green vegetation. The right hand ellipsis shows the location where we would ideally have placed this band, which is where the SPOT satellite based sensor has its SWIR band. The left hand ellipsis shows the location of our Skye sensor's fourth band. While the discrimination between green and dry pastures is not as distinct at ~ 1.03 um compared to ~ 1.7 um, based on these spectra we considered that there was enough potential for discrimination to go ahead with this wavelength choice of a fourth band.



REVIEWER #3: In page 18025 line 18. Please, add references to backup this statement

"The lower SWIR band of our multi-spectral sensors was also in the lower part of the SWIR range, which is not as responsive to dry vegetation as the longer SWIR bands."

AUTHOR RESPONSE: We could reference here both the Tucker (1980) article described in the methods section, and our examination of the spectral response of our sensors using the ASTER spectral library, both of which were the impetus for this statement in the manuscript. As mentioned in the above comment, Tucker (1980) recommends that the:"*1.55–1.75 µm region was the best-suited wavelength interval for satellite—platform remote sensing of plant canopy water status in the 0.7–2.5 µm region of the spectrum."*

REVIEWER #3: PAGE 18016 Methods. Same page lines 18-19. It is not clear to what complexity the authors refer to. In page 1817 lines 9-10. Incomplete sentence?

AUTHOR RESPONSE: We used ratio band indices as the Skye sensor model we used did not allow us to calculate calibrated reflectance. Without reflectance it is complicated to calculate more complex vegetation indices than simple band ratios and normalised band ratios. Skye documents an algebraic solution for calculating NDVI from the raw data that is equivalent to NDVI calculated from calibrated reflectance. The authors developed a similar algebraic solution for additional simple band ratio indices. This range of indices was considered sufficient for the pilot study development, and because other work had shown that simple ratios were effective for capturing seasonal differences in pasture biomass (e.g. Handcock et al., 2008). To clarify this we suggest extending this sentence to:

> "However, due to the algebraic complexity of calculating indices from this particular Skye sensor model (see the description in the paragraph below), we limited our index choice to simple ratios and normalized difference band ratios."

REVIEWER #3: PAGE 18016 Methods. Same page lines 14-17. Please review and rephrase.

AUTHOR RESPONSE: This can be rephrased in the revised manuscript.

REVIEWER #3: PAGE 18017 Same page lines 23-28 and next page lines 1-14. I would suggest to extend this part with additional explanations on the decisions taken.

AUTHOR RESPONSE: It would be possible to extend this section at the Editor's permission for paper length as, for brevity we have only given a small amount of detail here, and not repeated the material that is included in Table 2.

REVIEWER #3: PAGE 18017 For example, in table 2 it is not clear to me why there are to different criteria based on NDVI: NDVI < 0.1 and NDVI < 0.

AUTHOR RESPONSE: With regards to why the NDVI is tested twice for similar criteria in Table 2a and Table 2c, this is because the data filtering was designed to be explicit about why a value was excluded. We could just have said that we exclude all daytime NDVI of less than 0.1, as that would have captured both the < 0 and < 0.1 data ranges. However, by testing separately for each data range, we can distinguish in our data filtering between data that is excluded because of likely sensor noise (NDVI < 0.1) and those which are due to this being a tropical pasture (NDVI < 0). Also, if the filtering rules were moved to a different environment (e.g. ice, sand), the rule for NDVI < 0 might be adjusted , whereas the rule around data noise would always remain. This is why we have broken up Table 2 into different categories of error/filtering.

While we tried to make this clear by dividing up Table 2 into different categories of error/filtering, we could make this clearer in the revised manuscript by adding more detail as described above.

REVIEWER #3: Also, what authors consider a "brief" drop to zero or

AUTHOR RESPONSE: Any time when the noon RatioNS34 was zero that measurement was excluded as this is not an allowed value for our study environment. We specify "brief" here to indicate that that the drop to zero of RatioNS34 was an anomaly and not due to catastrophic sensor failure such as when we had a water incursion into the sensor case. We did not develop an automatic rule, but for any occasion where the RatioNS34 fell to zero we visually checked the database to see that the sensor values had returned to normal within one day, else the data was further inspected to see if this was an indicator of sensor failure or some other issue such as insects. An automatic rule could be easily developed where values of RatioNS34 of zero of 2 or more readings be flagged for more detailed analysis or a manual check. The word "brief" could be replaced with "1 day, unless manual examination indicated sensor failure")

REVIEWER #3: why the specific temporal period of the data source for filtering during daytime (12 to 13) was selected and why is it different from the time period of data used in the analysis (10 to 14).

AUTHOR RESPONSE: We differentiate in the first paragraph of section 2.7 between the time period that was chosen for calculating vegetation indices that are used for further analysis, and the second paragraph which talks about further data filtering during shorter periods which are detailed in Table 2.

The rational for choosing the period for calculating vegetation indices is currently explained in lines 15-17 as:

"multi-spectral sensor data, vegetation index values from the middle part of the day (10:00 to 14:00) were selected and the median value calculated to reduce noise due to small fluctuations in illumination."

The point the reviewer noted about further tests only having shorter time periods, for brevity is currently only detailed in Table 2. This could be expanded in the manuscript at the start of the second paragraph as follows:

"Other time periods were used for specific tests of the data, and for speed of processing, were at shorter time periods where this was considered suitable to the analysis, for example for the midnight or noon hour."

REVIEWER #3: In page 18018 lines 18-19. I agree that destructive sampling is not recommendable within the area under sensorts observation but, what about sampling in a nearby area? In the case of the fenced area could be even within the fence but out of the FOV. This should be considered an option unless authors assume that their measurements are so specific that can not be supposed representative of a larger area (which indeed would be not a good assumption).

AUTHOR RESPONSE: The goal of making measurements at the location under the sensors is to validate the sensor data at that location. Unlike more homogeneous temperate pastures, tropical pastures are naturally heterogeneous at the local scale, and the area around the sensors will be highly variable in both biomass and species composition. So, to sample at a nearby location could result in the field measures not being representative of the location that the sensor is observing. The goal of our pilot deployment was not to show that we could capture the heterogeneity of tropical pastures with just 2 sensors, but to show that it was viable to use such sensors for pasture monitoring. In the paper we discuss how ideally there would be a number of sensors deployed which capture the pasture heterogeneity.

REVIEWER #3: PAGE 18018 Methods. Same page line 23. Please, describe how these "standard photographs" where taken

AUTHOR RESPONSE: In the revised manuscript we suggest expanding on the description of the method to include some of the following points.

The photo standards are the industry standard for visual assessment of pasture status in this environment, and were created by the Queensland State Government:

"Standing feed can be estimated with the use of photo standards. The North Australian Grassland Fuel Guide (Johnson, 2002) and the Queensland Department of Primary Industries and Fisheries (QDPI&F) provide photo standards and the corresponding pasture yields on a CD-ROM."

Northern Territory. Dept. of Resources and Meat and Livestock Australia Cattle and land management best practices in the Top End region : 2011. Northern Territory Government. Dept. of Resources, Darwin, N.T, 2012.
http://www.nt.gov.au/d/Content/File/p/pi/topend/Best%20Practice%20Top%20End.pdf

The photo standards are provided on CD, and also on the following website, so that beef producers with similar pastures as those at the study site can have an industry standard for assessing pasture status. This is described on the following website as:

"Pasture photo standards are an essential tool in good grazing land management. They will help you develop pasture budgets and dry season business management plans. There are photo standards and corresponding pasture yields for many of Queensland's common pasture communities (based on region or pasture species)."
https://futurebeef.com.au/knowledge-centre/pastures-and-forage-crops/pasture-photo-standards/

Lees, Heather 2003, Pasture Photo Standards Agency for Food and Fibre Sciences, Beef, Department of Primary Industries, QLD.

The authors therefore selected these photo standards as the best option for non-destructive assessment of biomass at our study site because:

1. The goal of deploying sensors in this environment is to provide monitoring of grazing pastures used for beef production, and these are the industry standard photos
2. Examination of some of these photo standards will highlight how difficult these tropical pastures in northern Australian are to assess compared to the typically homogenous temperate pastures found in southern Australia. The recommended method for assessing pasture biomass in this environment is visual assessment, not destructive sampling.
3. The sensors monitored a fixed point on the ground for over 2 years, so destructive sampling of that footprint was not possible.
4. Taking additional samples in the around the sensor footprint is interesting to assess the heterogeneity of pastures in the area; however, those measurements should not be compared to the actual measurement under the footprint because of the pasture heterogeneity. In the southern temperate pastures we typically designed a hierarchical sampling scheme within an area, under the assumption that the southern pastures are homogeneous enough that the samples in an area should be representative of the whole area. In the topical pastures this is not the case, hence our field sampling design which measured the exact area under the sensor so that we could validate exactly what the sensor monitored.

If the authors were to design a new study, the same photo samples would be the best option of those available non-destructive methods, and the focus would still be on the area under the sensor. As we recommend future work to examine the spatial variability in these northern pastures to determine the best options for placing sensors and the number of sensors required, more extensive sampling of the pastures would be appropriate there.

Note that since we made our measurements, some of the pasture standards on the website were updated in 2013 with newer photos, but they are not substantially different.

REVIEWER #3: PAGE 18018 Same page line 24. Authors state in this paragraph that total biomass was not divided in the field between green and dry components because the technical difficulties, however, this is a key issue in this type of analysis (as it was further observed in the results). Trying to estimate total biomass using either proximal or remote sensing is a tricky question due to the contribution of the dry fraction to the signal. However, in this study, I assume that the estimation of the two fractions is a key subject from the operational perspective, so deeper elaboration on this issue is expected.

AUTHOR RESPONSE: Biomass was not split into dry/green as doing so for these heterogeneous tropical pastures is difficult when visual methods are used to assess biomass. The method does not lend itself to such sampling in the manner in which the area is visually assessed and then compared to photo standards, as there is no way of assessing the internal green/dry component that is not visible. We did, however, assess both the green and non-photosynthetically active vegetation (NPV) components in both 2 dimensions and 3 dimensions, as a way of capturing the green/dry component in lieu of destructive sampling. These measurements are detailed in the section following that identified (page 18019, lines 3-17).

REVIEWER #3: PAGE 18016 Same page lines 27-20 and page 18019 lines 1-12. Did the authors validate somehow this visual assessment? Visual estimates are commonly conditioned by the subjectivity of the interpreter Can the estimations be considered temporally consistent (same interpreter)?

AUTHOR RESPONSE: The visual estimates were not separately validated except to the photo standards, but they were carried out by trained field technicians. Based on metadata recorded with the measurements, almost 90% of the measurements were made by the same two research staff, with the other measurements made by 2 additional scientific staff, sometimes with the additional of the main two staff. Where more than one staff member took the measurements, data from that day was averaged. There were a total of 112 sets of field measurements prior to filtering and matching with the sensor data.

Post collection in the field the raw data was visually examined to see if there were any noteworthy differences between the data collected by the different operators, but no differences were discernible within the noise in the data. The measurements were therefore not controlled further for operator differences.

REVIEWER #3: PAGE 18016 Same page lines 27-20 and page 18019 lines 1-12. Did you tried to calibrate potential uncertainty in the estimations by comparing with direct measurements (destructive sampling)? This can be done in a nearby area (not the one observed by the sensors) and can be used to calibrate the indirect visual estimates.

AUTHOR RESPONSE: Although during the design of the field sampling the authors did consider whether destructive sampling should be made in the area around the sensor FOV, it was decided not to do so for the following reasons:

1. As described above, taking additional samples in the around the sensor footprint is interesting to assess the heterogeneity of pastures in the area; however, those measurements should not be compared to the actual measurement under the footprint. This is because of the heterogeneity of these northern tropical pastures. In the southern temperate pastures we commonly design a hierarchical sampling scheme within an area, under the assumption that the southern pastures are homogeneous enough that the samples in an area should be representative of a point. In the topical pastures this is not the case, hence our field sampling design which measured the exact area under the sensor so that we could validate exactly what the sensor monitored.
2. The goal of the present study is to develop methods that could be used for operational measurement on the farm, using methods available to beef producers. We therefore focused on the use of the visual standards provided by the Queensland state government, which themselves have been validated. We did develop a manual to be used by the research staff, to help ensure that the same methods were used over the long time frame of the study, and where possible had the same staff make the measurements.

We do recommend future work to examine the spatial variability in these northern pastures, to determine the best options for placing sensors and the number of sensors required. In this case more extensive sampling of the pastures would be appropriate.

REVIEWER #3: PAGE 18019 Methods. In page 18019 line 10. Check repetition?

AUTHOR RESPONSE: This suggested change can be made in the revised version.

REVIEWER #3: PAGE 18019 Methods. Same page line 18. Which area? 1x1 m? please specify

AUTHOR RESPONSE: The area is a 1m x 1m area under the sensor that was identified using pegs buried in the ground during sensor deployment. For brevity we do not detail this. This could be clarified by adding this to the first paragraph in the section, on Page 18018 line 15:

> "Visual observations of pasture biomass (weight of above ground vegetation dry matter (DM) per unit of area) for the sensors FOV were recorded by trained field staff at 2–3 week intervals during the study period, **in a 1 m x 1 m area under the sensors identified by small pegs hidden by the vegetation**."

REVIEWER #3: PAGE 18020 Methods. In page 18020 line 3. Use the acronym (VWC)

AUTHOR RESPONSE: Although we have used the term VWC in the description of the value that the sensor measures, as this is what is described in the Decagon documentation, the term "soil moisture" is used here as the short-hand for this measure to aid in clarity of interpreting the later tables of model results and reduced acronym use.

REVIEWER #3: PAGE 18020 Same page line 4. Yearly Julian date?

AUTHOR RESPONSE: We use the term "year day" as the value of interest in this southern hemisphere environment is the number of days since January 1 in the year. We did not use "Julian Day" as this refers to the days since noon Universal Time on January 1, 4713 BCE (see the NASA definition http://disc.gsfc.nasa.gov/julian_calendar.shtml).

REVIEWER #3: PAGE 18020 Methods. Same page lines 16-17. Please, specify the number of matching time points

AUTHOR RESPONSE: Of the original 112 sets of fields measurements there were 63 dates when the field measurements matched the field data. This can be clarified in the revised manuscript.

REVIEWER #3: PAGE 18020 Methods. Same page line 20. In fact there were not biophysical measurements at such but indirect estimations by visual analysis. Authors should clarify this statement

AUTHOR RESPONSE: The authors consider the observations of pasture biomass a biophysical measurement as the visual assessment records biomass as kg of dry matter per hectare. Additionally, we did directly measure other biophysical aspects of the pastures, such as the pasture height which was measured with a ruler and the soil moisture was measured directly.

This combination of measurement types could be refined in the text by replacing "biophysical measurements" by "biophysical measurements and other observations of the pasture and soil status"

REVIEWER #3: PAGE 18020 Same page lines 23-24. In my opinion this is one weak point in the statistical analysis. Collinearity analysis should be done as this could affect not only to the indices but also to other input data which has been combined in the regression as is the case of weather data, soil moisture and spectral indices. In fact, authors recognize that "soil moisture sensors. . . showed a strong correlation with the average of precipitation recorded at the CTA and TA stations" (page 18013).

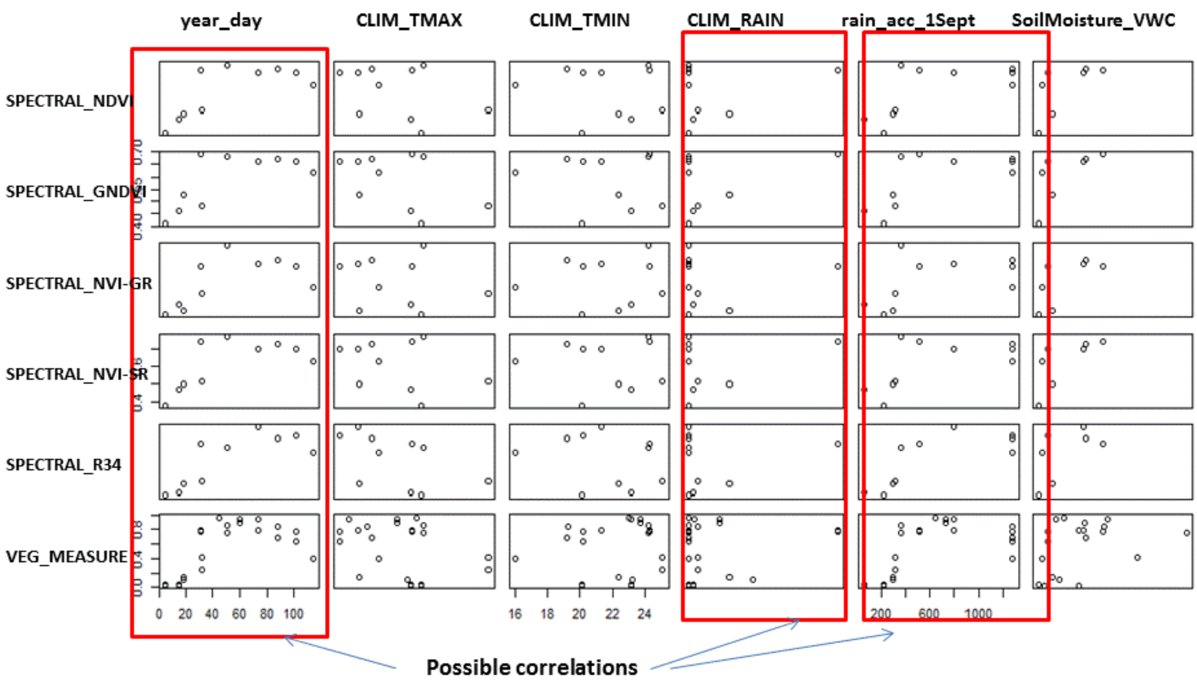
AUTHOR RESPONSE: The reviewer is correct in that there is likely to be a correlation between variables such as weather data, or between more than one vegetation indices. It was for avoid this problem of collinearity that the authors controlled the number and type of variables in the model development by:

1. Never using two sensor variables in a model
2. Never combine two climate/environment variables such as soil moisture and precipitation.
3. For the 2-variable models, combining only one sensor variable with one climate/environment variable.
4. Some variables which we measured in the field were not included in the paper as they were found to be too highly correlated with other variables and so would not add to the analysis.

As part of model development we did extensive exploratory data analysis of many combinations of variables, to examine the relationships among variables and reinforce the above variable selections.

The aim of this analysis was to test the strength of relationships between the field and sensor data, as part of the pilot project. If we could show that after all of the data quality control that it would still be possible to develop relationships between the sensor and field data, then this would be the basis for further work to extend the relationships.

As part of addressing the reviewer's comment we re-visited our initial exploratory analysis including the scatterplots and correlation matrix (see below).



All statistical analyses in this project were done in R, with the Pearson product-moment correlation coefficient calculated for each pair of variables used in the 2-variable models as follows:

	year_day	TMAX	TMIN	RAIN	rain_acc_1Sept	SoilMoisture_VWC
NDVI	0.69	-0.42	0.02	0.19	0.71	0.66
GNDVI	0.65	-0.43	0.08	0.28	0.66	0.67
NVI.GR	0.62	-0.36	0.07	0.14	0.67	0.69
NVI.SR	0.64	-0.39	0.08	0.26	0.67	0.70
R34	0.73	-0.49	-0.03	0.03	0.72	0.58
VEG_MEASURE	0.53	-0.31	0.19	0.26	0.55	0.77

Assuming a threshold off of 33%, these correlation coefficients indicate that four of the climate/environment variables have medium-high correlations with the spectral variables, and that only the minimum-temperature and the rainfall should be included in 2-variable models. This is an oversight in our data analysis, which initially set the threshold much lower, and we thank the reviewer for suggesting that we re-visit these earlier analyses.

In the revised manuscript we therefore will remove the subset of the 2-variable models that show this high correlation, and change the results discussion, tables, etc. to reflect this. This change will not impact on the most important relationships discussed in the paper, which are the single variable relationships between the sensor and field measurements.

REVIEWER #3: PAGE 18020 Same page lines 23-24. Also, additional information should be included on the statistical method applied, especially on statistical indicators for model evaluation that are later analyzed on the results section (tables 3 to 5).

AUTHOR RESPONSE: See above comments around the collinearity discussion, which also indicate some of the additional model development and exploratory data analysis that was made as part of the study. Much of this analysis was not discussed in the paper due to brevity, and some of this detail has been made instead in these comments. We would be glad to include some of the detail around model choices in the paper, rather than encapsulating in the brief descriptions that we currently have.

REVIEWER #3: PAGE 18021 In page 18021 lines 4-6. This is also a key point in the statistical analysis. Authors state that the standard split-sample validation was not performed due to the insufficient data. However alternative techniques as bootstrapping could be applied to analyze generalizability and reproducibility of the models. See Richter, K; Atzberger, C; Hank, TB; Mauser, W(2012): Derivation of biophysical variables from Earth observation data: validation and statistical measures.J APPL REMOTE SENS. 2012; 6

AUTHOR RESPONSE: We agree with the reviewer that a bootstrapping approach would be an option for performing some validation, given our limited dataset.

In the present study we did not implement such a method as we were concerned about the sample size of the dataset once we excluded all but the wet-season data. Of the 63 times where we had matching field data and at sensor data from either the multispectral sensors or the digital cameras, only 23 remain when we look at just data from the wet season. Of these, only 10 matching data pairs remain where there is a matching multispectral measurement. We considered a sample of 10 to be too small for bootstrapping or the "leave one out" category of methods.

However, we would be interesting to do the bootstrapping analysis if the Editor thought that it would be of sufficient interest to readers even given the very limited sample size. Either way, we suggest that in the revised manuscript we briefly mention these options and decisions, and reference the excellent paper that the reviewer suggested.

When we initially planned the study, the expected database size of 112 sets of field measurements would have been sufficient for a bootstrap analysis, but given that the exploratory data analysis in both visual and preliminary statistical form showed that the relationships in the dry season were uniformly poor, we had to exclude the dry-season data from our analysis. Future work could maximise the number of matching data points by focussing field sampling efforts just on the wet season. This would provide enough data for validation and calibration, and/or for one of these alternative methods.

REVIEWER #3: In page 18022 lines 4-8. But this is not related with the distance between the nodes, is a consequence of the differences in grazing activity (fenced vs not fenced), isntt it? This should be clarified in the text.

AUTHOR RESPONSE: The fencing of one node to exclude grazing was made to extend the range of vegetation height and therefore the range in the calibration data, but due to the limited feed availability in the paddocks these grazing events had negligible impact on the on the grazing. Our intention in describing the nodes as being only 200 m apart was used as a short hand for saying that the vegetation at the two nodes had similar (if still high) heterogeneity. This description can be revised in the revised manuscript.

REVIEWER #3: PAGE 18022 Results. Same page line 16. Lost periods should be expressed in relative terms as in section 3.1

AUTHOR RESPONSE: This can be clarified in the revised version: "13 days of data from the unfenced node (3%), and 10 days of data from the fenced node (2%)"

REVIEWER #3: PAGE 18022 Results. Same page line 18. It would be very useful to include in the images the approximate area observed by the Skye sensor FOV.

AUTHOR RESPONSE: These images could be annotated to indicate these approximate area observed by the multispectral sensors. If this was added to every image it would, however, add clutter to the time-series and distract from the figure's goal of illustrating what the seasonal progression of vegetation looks like in this environment. A compromise would be to add this annotation to only the first figure in the series. We will leave this decision to the editor.

REVIEWER #3:. In page 18023 lines 10-11. This residual standard error (RSE) is equivalent to the most commonly used RMSE? I would recommend using a relative measurement such as the NRMSE or RRSME (see Ritcher et al 2012) which is more easily comparable.

AUTHOR RESPONSE: The difference between the RSE and the RMSE is the denominator. The residual standard error (RSE) is the square root of the sum of the squared residuals / degrees of freedom, whereas the Root Mean Squared Error (RMSE) is the square root of the sum of the squared residuals / n :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(p_i - a_i)^2}{n}}$$

The paper that the reviewer recommends(Richter et al., 2012) highlights the lack of standardisation of statistical terminology. The NRMSE and RRSME suggested would both be good alternatives for assessing error, and can be addressed during paper revisions.

REVIEWER #3: PAGE 18023. Same page lines 25-26. But, is this really useful for farmers? In page 18009 authors mention that farmers need to assess, at the end of the wet season, how much green feed remains in the paddocks. Therefore I assume that the most useful information, from an operational perspective, would be to provide an accurate estimation of the green and dry fractions. These seem not to be possible if the analysis is restricted to the wet season, where green fraction dominates

AUTHOR RESPONSE: The green fraction makes up most of the biomass during the wet season and as the season finishes, rainfall decreases, soil moisture decreases and green biomass dies to become dry forage. Therefore, accumulation of green biomass throughout the wet season until its end will also be an indication of the amount of forage at the end of the dry season. Farmers need to monitor accumulation of biomass throughout the wet season to ensure they have the amount of forage needed at the end of the wet season which will carry their cattle through to the end of the dry season.

REVIEWER #3: PAGE 18025. Same page lines 21-24. Please, rephrase

AUTHOR RESPONSE: This suggested change can be made in the revised version.

REVIEWER #3: PAGE 18025. Same page line 25. Please, add discussion here on what authors expect a non-pilot study could add apart from having in situ weather data.

AUTHOR RESPONSE: The form of a non-pilot study would depend on the purposes that a future deployment is for. For example whether it is to deliver operational data to farmers, or to validate satellite data, to test the design of sampling schemes using many low-cost sensors, to use the sensors for monitoring an area for degradation etc. There are also many technical directions such a project could go, such as sending image data back across a sensor network, processing data at the node, etc. An in-situ weather station may not be a part of a future deployment depending on its needs, as the soil moisture data may be considered sufficient. We have tried to address these problems individually in detail in the discussion section so that the information can be used in the design of future deployments. We currently summarise the possible directions of a future study in the conclusions section, which should hopefully be sufficient, but could expand this if the Editor recommends.

REVIEWER #3: In page 18026 lines 24-26. Please avoid repetition “available”

AUTHOR RESPONSE: This suggested change can be made in the revised version.

REVIEWER #3: Results. In page 18027 lines 3-4. Why this type of filtering was not implemented in this pilot study?

AUTHOR RESPONSE: The instances of surface water were minimal (9 days at the fenced node and 20 days at the unfenced node), so the data was simply excluded manually using the digital images, particularly as they fell during the period when there was water incursion into the sensor housing and the data was begin masked anyway . Automated filtering was not implemented because it was considered a secondary consideration to the main goals of the study which was to monitor the vegetation. The number of days of surface water was also thought to be too limited a dataset to develop and test robust filtering rules, particularly as the sensor data at that time was suspect. We wanted to mention how it might be done however, in a more robust sensor deployment.

REVIEWER #3: PAGE 18027. Same page lines 13-14. How authors propose model validation? Spatial/temporal strategy? It would be interesting to include further discussion on that key question.

AUTHOR RESPONSE: We briefly touch on how the model could be calibrated at different studies sites in the sentences following those indicated by the reviewer:

> "*If further studies do not show consistent relationships between sites and years, one option for calibration would be to have the farmer performing a controlled set of calibration measurements once or twice during the growing season to calibrate a particular sensor deployment*."

These comments could be extended to reinforce why we made this suggestion, for example for by adding:

> "While having to make some measurements is an additional time requirement, by being made at the same location as the deployed sensors such measurements would alleviate the cost of gathering the large amount of calibration data required to develop models that would be robust for different geographical locations and different weather conditions between years, and changes in the calibration of the physical sensor over time. Alternatively, the time-series of vegetation index data from the sensors could be used without calibration to a quantitative value, which would still provide data to indicate sudden changes in vegetation growth"

REVIEWER #3: PAGE 18028. In page 18028 line 13. It would be interesting to know what was the problem related with sensor failure.

AUTHOR RESPONSE: The sensor that failed completely had water incursion into the sensor enclosure. Examination of the sensors and discussion with the manufacturer suggested that moisture entered the sensor enclosure at the point where the wiring attached to the sensor exclosure. This was despite sealant being applied to the connection and regular monitoring . Given that we had a spare sensor that could be used as a replacement the decision was made to swap the sensors out to ensure continuity of data collection while the sensor was returned to the manufacturer for examination.

REVIEWER #3: PAGE 18029. In page 18029 line 6. This is an interesting issue that could be also explored and demonstrated with the dataset provided by this pilot study.

AUTHOR RESPONSE: The suitability of the data capture depends on the purpose for which the sensors are deployed. In our pilot study we used data from Day-time (12:00 to 13:00) and Night-time (00:00 to 01:00) for quality control, camera images nearest in time to 12:00 were selected, but other images were used for visual examination. The topic could be addressed by adding the following point to the manuscript:

> "For example, we showed that a single image selected in the middle of the day was sufficient for seasonal monitoring, but that camera images from other times of the day were also useful for investigating unexpected data from the other sensors".

REVIEWER #3: PAGE 18029. Same page line 27. Please explain what authors consider "sufficiently sampled". The definition of the spatial and temporal sampling strategies is a key point in this type of approaches so it is important to discuss here.

AUTHOR RESPONSE: The choice of sampling scheme would vary depending upon the heterogeneity of the pasture and the accuracy required for farm management for their particular the purpose. We touch on this at the end of the sentence by saying "*so that the expected spatial variability in the paddock is covered to enable farm-management decision to be made at critical points in the season*.", and also in the discussion in the paragraphs that follow this sentence. However, this point could be clarified by expanding the shorthand-wording that we have used of "sufficiently sampled" to:

"Sampled at a sufficient number of points, and those points arranged in such a way so as to capture the spatial pattern in the particular landscape. This includes, for example, whether the spatial pattern in the pastures is relatively stable, as is more common in temperate pastures, or is more clumped and heterogeneous as in common in tropical pastures, as well as changes to the spatial heterogeneity of the pastures due to pasture management such as re-seeding. The assessment of landscape spatial pattern at multiple scales is a broad topic, but good overview can be found in McCoy (2005), and a more detailed example in Chen et al (2012)".

- McCoy, R. M.: Field methods in remote sensing, Book, Whole, Guilford Press, New York, 2005.
- Chen, Baozhang, Nicholas C. Coops, Dongjie Fu, et al. 'Characterizing Spatial Representativeness of Flux Tower Eddy-Covariance Measurements Across the Canadian Carbon Program Network using Remote Sensing and Footprint Analysis', Remote Sensing of Environment, vol. 124/(2012), pp. 742-755.

REVIEWER #3: PAGE 18031 Conclusions. In page 18031 line25. Please rephrase to avoid repletion "limiting data acquisition"

AUTHOR RESPONSE: This suggested change can be made in the revised version.

REVIEWER #3: Figures. In figures 2 and 3. Remove NDVI range in the title of y axis (not necessary).

AUTHOR RESPONSE: This can be changed in the figure assuming that the editor confirms that this is standard formatting style for unit less values in Biogeosciences, as the formatting style for unit less measures differs between journals.

REVIEWER #3: Figures. In figure 3. Remove horizontal axis for NDVI = 0. The x axis should start the first day with spectral data available and finish also with the last day of data available (ot before and after).

AUTHOR RESPONSE: These changes can be made in the revised version.

REVIEWER #3: Figures. In figure 4. The image on the left looks blurry. A better quality image would be recommendable

AUTHOR RESPONSE: Unfortunately this is the best quality image that we had showing the mid wasps.