Response to reviewers

Response to comments and suggestions from Reviewer 1

General comments

The paper presents a relevant contribution to precision agriculture by coupling NDVI and EMI for management zones' delineation. The methodology is generally sound, especially the use of the SOM and MCASD for cluster optimisation. The study presents a robust workflow that could inform both research and practice. However, some aspects need clarification to improve generalisability and interpretability.

We thank the reviewer for the positive assessment, and are happy to provide the requested clarifications.

Specific Comments

- Lines 64-122: The review of EMI and NDVI is largely descriptive. It would be stronger if the authors synthesised how of the previous studies succeed or failed in integrating these data types. I suggest adding a short synthesis paragraph summarising what's missing in prior work and how this study fills the gap.

We thank the reviewer for this helpful suggestion. In response, we have now included the following statement in section 1 (lines 118-128):

"In summary, while previous studies have made important contributions towards integrating EMI and NDVI data for management zone delineation (Corwin and Scudiero, 2019; Ciampalini et al., 2015), the results have been highly dependent on sensor resolution, data timing, and local soil-plant interactions. Some studies demonstrated that EMI alone offers strong insights into soil structure and moisture patterns, and suggested that crop-level responses captured by NDVI can be inconsistent due to seasonal and environmental variability. Others highlighted the value of combining datasets but faced limitations in spatial resolution, ground-truth validation, or field-specific conditions that restricted the precision of zone delineation. This study builds on these efforts by combining high-resolution EMI and NDVI data within a harmonized framework, applying consistent normalization, and validating the resulting zones with multi-year yield data and dense soil sampling."

- Lines 304- 309: the use of min-max scaling prior to clustering is appropriate for ensuring feature comparability. However, the authors should briefly justify this choice over alternatives (e.g., standardisation, robust scaling), especially given the potential presence of outliers in EMI and NDVI data. Min-max is sensitive to extreme values, which may distort the input space and affect cluster geometry in SOM.

We thank the reviewer for pointing out this important consideration. In our study, EMI data were already filtered to remove outliers from a variety of sources. In particular, the combination of min-max filtering, histogram filtering, and ECa variation filtering effectively remove outliers from the EMI data, as shown in previous research. We therefore are confident that the resulting distribution of ECa values, combined with the use of z-transform normalization, is appropriate for a min-max scaling. Similarly, NDVI maps were pre-processed by PlanetScope to remove atmospheric artifacts, and we manually excluded data from periods with low vegetation signal (lines 452-457). Moreover, the extent of the area and the amount of pixels in the NDVI images assures that the distribution is free from outliers that would affect the min-max scaling.

Nonetheless, we understand that this may not be the case in other areas or when different data sources are used. Thus, we now addressed these topics in section 3.5, where the new text reads:

"Although min-max scaling was suitable in this study due to the relatively smooth and filtered input data, it is known to be sensitive to outliers and data range extremes. In datasets with greater variability or different preprocessing methods, alternative scaling approaches such as standardization or robust scaling could be more appropriate. Future studies should assess the impact of different normalization strategies on clustering results, especially in settings with noisier or unfiltered sensor data."

- Lines 431-351: the authors perform 100 SOM runs per candidate cluster number and use the MCASD to select the optimal k. While this addresses compactness, there is no assessment of cluster stability. Please clarify whether variability across SOM runs was quantified (ARI or some cluster overlap metrics).

We thank the reviewer for this valuable comment. While we did not explicitly compute clustering overlap metrics such as the Adjusted Rand Index (ARI), our approach used the Multi-Cluster Average Standard Deviation (MCASD) inherently reflects variability across SOM runs. Specifically, MCASD quantifies the stability of cluster centers by averaging their standard deviation over multiple iterations. During preliminary testing, we observed that most datasets stabilized in terms of variability between 70 and 80 iterations. To ensure consistency and reproducibility, we adopted 100 runs per cluster number. This approach provided a reliable means to assess both compactness and relative stability of clusters in a computationally efficient manner. We have clarified this in the manuscript and added a note in the Limitations section (Section 3.5) to suggest the use of additional stability metrics like ARI in future work. The new text reads:

"While cluster variability was addressed using the Multi-Cluster Average Standard Deviation (MCASD) across 100 SOM runs, future studies may benefit from incorporating additional stability metrics such as the Adjusted Rand Index (ARI) or cluster overlap measures to better assess classification consistency."

- To enhance the clarity of the manuscript. The authors should consider including a workflow diagram summarizing the complete methodology.

We thank the reviewer for this helpful suggestion. To enhance clarity, we have added a workflow diagram (Figure 2) in Section 2.2 that summarizes the complete methodology, including the classification and validation steps. The diagram visually outlines the integration of EMI and NDVI data, the clustering process using SOMs and MCASD, and the post-hoc validation using yield and soil data.

The overall methodology of this study, including data, processing steps, and validation is summarized in Figure 2. This flowchart highlights the role of EMI and NDVI datasets in clustering process and the use of multi-year yield maps and soil samples for validation and refinement of the resulting management zones.

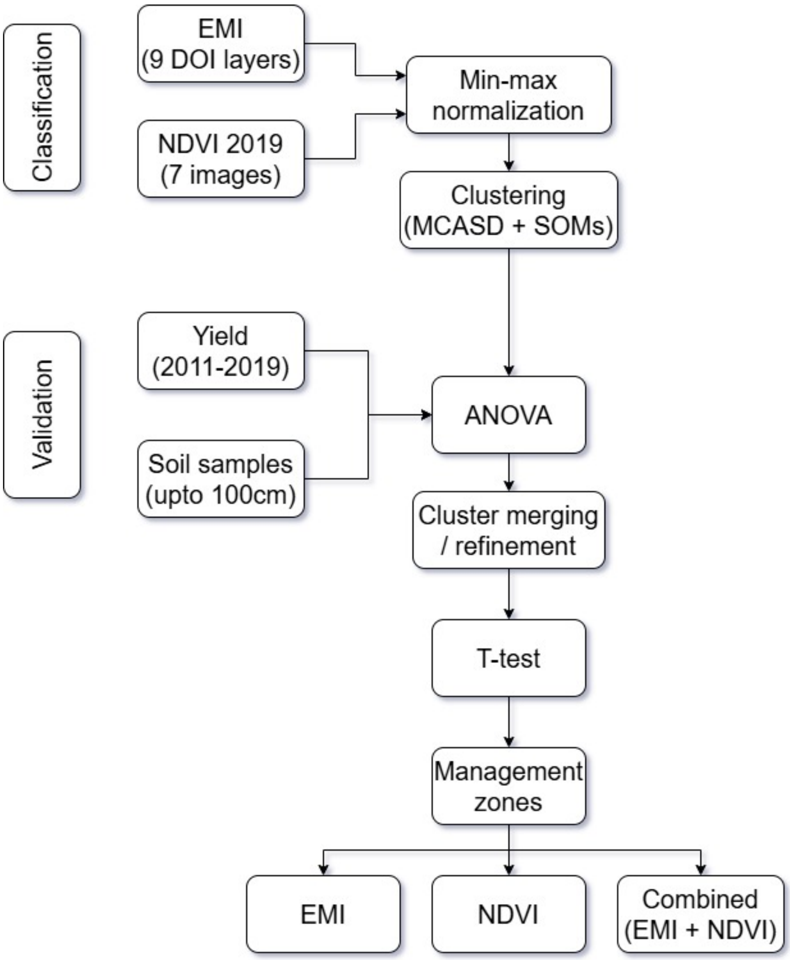

Figure 2. Workflow diagram showing the integration of proximal (EMI) and remote sensing (NDVI) data for unsupervised clustering using MCASD and SOMs. Yield and soil datasets were used for post-hoc validation and refinement of management zones.

- Lines 93-96: while NDVI is common vegetation index, it is well-known to saturate under high biomass or dense canopy conditions, which may limit its ability to capture within field variability during peak crop growth. The authors should justify why NDVI was selected over alternatives such as EVI or SAVI.

We thank the reviewer for the insightful comment. We acknowledge that NDVI can exhibit saturation under high biomass or dense canopy conditions, which may limit its sensitivity during peak growth. However, we used NDVI as: a) it can directly and reliably be derived from the PlanetScope sensor as well as from many other sensors (e.g. satellite-, aerial- and drone-based), b) the focus of our study was on capturing relative spatial variability within the field, not absolute vegetation productivity, and c) NDVI remains a widely accepted, validated, and simple index for evaluating vegetation vigour across phenological stages. In fact, other indices like EVI and SAVI can require specific calibration parameters (e.g., soil brightness correction factor or coefficients tied to aerosol resistance), which were not feasible to constrain accurately within our satellite dataset and field setting. We thus preferred to use NDVI, which does not require additional computation or calibration. We think that this makes for a simpler, ready to use, and transferrable approach. To avoid extending an already long manuscript, we would prefer to not provide additional justification in the manuscript.

- Lines 370-395: The initial presentation of yield maps provides useful spatial context. However, since the 2012 and 2013 data are acknowledged to be lower in quality, the authors should discuss whether these data were weighted differently or excluded from statistical validation to avoid introducing bias in zone validation.

We thank the reviewer for this important point. The 2012 and 2013 yield data were presented because they showed relevant spatial trends, despite lower data quality. To avoid introducing bias, these years were not weighted differently in the statistical validation. Instead, we relied on multi-year averages and year-by-year comparisons to assess the robustness of zone delineation. This clarification has now been added to the end of the yield data subsection (Lines 393–397). The new text reads:

"... they were retained for spatial context as they still exhibited consistent patterns with other years. These years were not weighted differently during validation analyses, and the potential influence of this lack of weighting was mitigated by evaluating multi-year trends and conducting year-by-year comparisons in the validation stage (see Section 3.4)."

- Given the spatial nature of EMI and NDVI data and the use of kriging interpolation, spatial autocorrelation is likely present in the dataset. While the current clustering is sound, the authors may consider briefly acknowledging the presence of spatial structure and its potential influence on post-hoc tests.

We thank the reviewer for this comment. We agree that kriging interpolation introduces spatial structure in the EMI and NDVI datasets, which can influence the assumptions underlying post-hoc statistical tests such as ANOVA and t-tests. While we did not explicitly correct for spatial autocorrelation, we believe its impact was mitigated through the use of multi-year yield data and non-interpolated soil sampling in the validation process. We have now included an explicit acknowledgment of this point in the Limitations section. The new text reads:

"This may influence statistical outcomes or lead to less spatially coherent clusters in some cases. Additionally, the use of kriging interpolation for EMI and NDVI datasets introduces spatial structure that may further affect the assumptions underlying post-hoc statistical tests."