Response to Reviewer 1

How to cite. Ubeda-Trujillo, I., Werner, M., Bertini, C., Coenders-Gerrits, M., and Jewitt, G.: Detection of agricultural flash droughts using impact-informed thresholds, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2026-1010, 2026.

We thank the Reviewer for the detailed questions, insightful comments, and time invested in reviewing the manuscript. These suggestions have helped us improve the clarity of the paper. The reviewer's comments and the authors' responses are provided below, with reviewer comments shown in green and responses shown in blue.

Major Comments

1. I felt the manuscript needs a language editing to improve the overall readability for some sentences and paragraphs. There were typological errors and some sentences that needs grammatical correction for clarity and coherency. Consider also the language preference for the journal (English guidelines) as well as the overall structuring of the manuscript (Submission).

We thank the reviewer for the comment. The manuscript will be carefully revised to improve readability, clarity, and coherence. We corrected the typographical errors throughout the text and revised sentences and paragraphs for improved flow and consistency with language guidelines of the journal.

2. In the abstract, it introduces the problem, aim, and results. This can be further improved by highlighting the main results in line to the aim. In the current text, the results mainly presented the limitation of the literature-based approach instead of how well the proposed approach improve the detection of AFD as compared with literature-based approach. Give an emphasis on the skill of the new approach in detecting AFD over literature-based approach.

We agree with the reviewer that the abstract placed greater emphasis on the limitations of literature-based thresholds rather than on the performance of the proposed approach. To address this, we have revised how the results are presented in the abstract to more explicitly highlight the performance of the calibrated impact-informed thresholds relative to the literature-based thresholds.

3. Method section can be further improved by structuring a coherent flow of the processes in conjunction to the objectives presented. A confusion on the dataset described in section 2.2 where MODIS MYD16A3GF having annual temporal resolution was later describe in succeeding text of using MODIS data with an 8-day resolution specifically on calculating SESR. Moreover, a need for a more detailed description of the specification of the datasets being used.

We thank the reviewer for this comment. We agree that the methodological workflow can be presented more clearly in relation to the study sub-objectives answered in this study. In the revised version we propose to a minor restructuring of the method section and have renamed the sub-headers of the section to increase the coherence and align the steps directly with the sub-objectives. In line with this, the spatial and temporal sampling design section will be incorporated directly into Results Section 3.1 instead of being presented separately in Section 2.3.1.

We also agree that the description of the MODIS dataset requires clarification. The previous version incorrectly referred to the annual MYD16A**3**GF product instead of the MYD16A**2**GF, which provides 8-day evapotranspiration data at 500 m x 500 m spatial resolution and was used for SESR calculation. This mistake will be corrected, and we will include further details on dataset characteristics, temporal and spatial resolution, and the role of each dataset within the study.

4. Methods sections 2.2, 2.3, if possible, provide flowcharts for ease visualization and reduce the text intensive description of the procedure or processes done.

We thank the reviewer for this suggestion. We agree that Sections 2.2 and 2.3 contain extensive procedural descriptions. In section 2.2 and 2.3, rather than introducing a flowchart, in the revised version we propose streamlining the description of the preprocessing procedures by focusing on the main processing steps rather than explaining the rationale behind each step, while retaining the corresponding methodological citations for readers interested in further details.

5. In section 2.3.2, what do you mean by the 3km sampling interval? It is not clear to me how the sampling method is done. Does it mean you selected a 3km x 3km grid for each soil texture? the 388 timeseries is also not clear which time scale it is referring to? I am confused with the data being processed because the MODIS MYD16A3GF has annual temporal resolution and here you are mentioning an 8-day SESR values? In addition, why do you need to do sampling? is it not possible to process all the pixels covered by each soil texture? was the data too large that takes too long to perform the computations?

We thank the reviewer for pointing this out. The reported 3 km value does not refer to a fixed 3 km × 3 km grid. Rather, it represents the approximate average spacing between sampled locations derived from the total area under study and the final number of samples. The average spacing was estimated as $d = \sqrt{\frac{A}{N}}$. where d is the approximate average spacing between sampling locations, A is the total area (km²) of the study area, and N is the total number of sampling locations. This relationship provides the expected average distance between sampling locations, assuming an approximately uniform spatial distribution within rainfed areas across the soil texture classes of the study area.

Processing all pixels would have been technically possible, but it would have substantially increased computational demand. Therefore, a stratified sampling approach was adopted. Soil texture classes were treated as strata, and sampling locations were allocated proportionally to the area occupied by each soil texture class while maintaining a minimum of 30 locations per stratum. A minimum of 30 locations per soil texture class was selected as a practical representation constraint in the stratified sampling design during calibration. The 388-time series therefore refer to 388 sample locations, each represented by an 8-day SESR time series for the period under analysis for February to August during the 2000-2024 study period. As mentioned above, the spatial and temporal sampling design will no longer be presented as a standalone methodological step. Instead, its results will be incorporated into Results Section 3.1, where they directly support the corresponding study sub-objective.

Regarding MODIS MYD16A3GF, we agree that this was confusing in the previous version. As noted above, the manuscript incorrectly referred to the annual MYD16A**3**GF product. In the revised version, we will correct this by specifying that the study used the MYD16A**2**GF product, which provides 8-day evapotranspiration and potential evapotranspiration data at 500 m x 500 m spatial resolution. Therefore, the 8-day SESR time series were derived from the MYD16A2GF product and not from the annual MYD16A3GF product.

6. In section 2.3.2, what is the relevance on why you consider the criteria of having VHI below 0.35 at the end of cropping season as indication of drought years with crop losses? why not use the VHI on the first month of the growing season, or the VHI at the 3 month of the growing season which coincides with the flowering and yield formation stage for maize and the most sensitive stages to drought which could cause yield reduction or crop damage?

We thank the reviewer for pointing this out. We agree that drought occurring during early growth or flowering stages can substantially affect crop development and final yields. However, our objective was not to identify periods of crop sensitivity during the season, but rather to construct a benchmark distinguishing years with and without drought-induced crop losses for threshold calibration. For this purpose, we adopted the criterion proposed by Fuganti et al. (2020), where VHI values below 0.35 at the end of the first cropping cycle in Nicaragua (August) are interpreted as crop losses in the field. In contrast, VHI conditions earlier in the season are primarily used as indicators of vegetation stress and early warning conditions, which do not necessarily imply actual crop losses at the end of the season. To clarify this rationale, we will add this explanation.

7. In section 2.3.3, is the first paragraph describing the literature-based AFD detection? If so, I think the title needs to add a term like 'literature-based AFD detection' just to showcase the method. It is also better if a more detailed description on how this approach (literature-based) is applied to detect an AFD for a reader who is not familiar with the method. At first reading, I could not understand how the method works. Moreover, a separate section to discuss the proposed enhancement to the existing method is suggested. In this way, as a reader I can separately compare the two approaches.

We thank the reviewer for this suggestion. Yes, the first paragraph of Section 2.3.3 describes the literature-based SESR flash drought detection method proposed by Christian et al., (2019). This study does not introduce a new flash drought detection. Instead, it retains the SESR-based flash drought method of Christian et al., 2019, including the rapid intensification thresholds criteria, and modifies only the drought onset threshold criterion. As stated in lines 265–266, “This study refines this framework by replacing the fixed 20th percentile threshold with a calibrated threshold specific to crop management periods and soil texture”.

In the revised manuscript, we will clarify this distinction by explicitly stating that the literature-based approach refers to the original SESR flash drought detection method of Christian et al., 2019, in which the drought onset threshold is defined using a fixed SESR percentile and is not calibrated against local agricultural drought impacts.

8. IN SECTION 2.3.6, how do you evaluate the performance of the calibrated threshold and the performance of the literature-based approach? what are the metrics you measured to determine whether an improvement is achieved by the calibrated SESR Threshold? what does it mean by V1 =1 or V1 = 0, does it mean a hit or miss? For example, for year 2014, if it is V1=1 then it is a hit meaning it predicted or detected the occurrence of AFD? what is crop loss condition = 1 means?

We thank the reviewer for this comment. V1 and V2 are binary validation metrics used to assess whether events detected using the fixed or calibrated SESR drought onset thresholds together with the same rapid intensification criteria coincided with the two agricultural impact benchmarks. A value of 1 indicates coincidence between a detected event and the corresponding benchmark (hit), whereas a value of 0 indicates no coincidence (miss). Specifically, V1 = 1 indicates coincidence with the VHI-derived agricultural impact benchmark, whereas V2 = 1 indicates coincidence with documented drought impacts reported in the benchmark literature.

Because the study compares the same detection method under two different drought onset threshold approaches (fixed and calibrated threshold), improvement was assessed by comparing the agreement of detected events with the two agricultural impact benchmarks. The calibrated SESR threshold was considered to improve performance when it showed greater agreement with the agricultural impact benchmarks than the fixed literature-based threshold.

9. In section 3.1, for the results, since literature-based approach have already used the SESR as the parameter in detecting AFD, why do you need to correlate SESR with VHI. A need to expound the relevance of this results in connection to the overall aim of the study. In addition, I also don't see results summarizing the identified years having drought with crop losses and without crop losses used in the calibration and validation. For transparency of the data used, please include the results maybe as supplemental tables.

We thank the reviewer for this observation. Indeed, SESR has already been widely applied in the literature for flash drought detection, and the purpose of Section 3.1 was to assess whether evaporative stress, responded differently between years with and without drought-induced crop losses under the dryland conditions of the study area.

We correlated SESR with VHI because VHI was used as an agricultural impact benchmark. The objective was to determine whether stronger evaporative stress signals captured by SESR were associated with poorer vegetation conditions and crop losses. This objective was particularly important because, in dryland environments, high evaporative demand and chronically low soil moisture make evaporative stress part of the normal background conditions. Under these conditions, it cannot be assumed that literature-based SESR thresholds consistently distinguish years with and without drought-induced crop losses across crop management periods and soil textures. Therefore, before calibrating impact-informed SESR thresholds, it was necessary to assess whether SESR differentiated years with and without agricultural impacts. If no such relationship existed, the subsequent calibration of SESR thresholds against agricultural impact benchmarks would have lacked a sufficient basis.

Regarding the identification of years with and without drought-induced crop losses used during calibration and validation, Figure 3 was included as an illustrative example of the benchmark construction procedure using both the VHI-based and documented-impact benchmarks at a single representative sample location. The purpose of the figure is to demonstrate how years with drought-induced crop losses were identified. Because benchmark datasets were constructed independently for each of the 388 sampled locations and soil texture classes, it is not practical to present equivalent figures for all sampled locations. We will clarify this point in the manuscript and indicate in the Data Availability section that the processed benchmark datasets used for calibration and validation can be shared upon request.

10. In section 3.2, Performance was presented however, it is also important to showcase the calibrated threshold as it was presented as a supplemental material Fig. S3. A section should be allotted for the presentation of the variation of the calibrated optimal thresholds prior to section 3.2. In addition, the context where SESR threshold being calibrated for the different 'crop management period' was not implemented, instead the results showed a calibrated SESR thresholds at monthly basis and not on the time frame for each crop management period. Furthermore, the calibrated SESR threshold was not clearly describe how it was selected among the incremental runs of maximizing. It was also confusing when I relate it in section 3.3 where it presents plots of optimal SESR thresholds. There were several optimal SESR derived but how did you select the best SESR presented in Fig. 6 among these SESR?

We thank the reviewer for this comment. The calibrated SESR thresholds are already presented in Figure 6. In each panel, the solid vertical line denotes the optimal calibrated threshold ($\Theta_{t,k}$) obtained through the Fβ maximisation procedure for a specific crop management period and soil texture class. Figure 6 therefore presents both the calibrated threshold values and the behaviour of the associated performance metrics across the full threshold range. Figure S3 serves a different purpose. Rather than presenting additional calibrated thresholds, it illustrates the variability of the optimal calibrated thresholds obtained from the bootstrap procedure and therefore provides information on threshold uncertainty. Subsequently, Figure 8 summarises this uncertainty through the CI95 metric and identifies which optimal calibrated thresholds satisfy the predefined performance and uncertainty criteria required to be retained as skilful calibrated thresholds for validation. To avoid confusion, we propose clarifying in the revised manuscript the different roles of Figures 6, 8, and S3, namely the presentation of optimal calibrated thresholds, their uncertainty, and the identification of skilful calibrated thresholds, respectively.

As described in lines 127–129 in the manuscript, the crop management calendar comprises the pre-sowing, sowing, growing and harvesting periods. In this approach, April corresponds to the pre-sowing period, May to the sowing period, and June–July to the growing period, whereas harvesting (August) was not included in the threshold calibration because it represents the impact assessment stage. The threshold optimisation was conducted using the original 8-day SESR observations from February to July, yielding one optimal threshold for each soil texture–month combination. The resulting monthly thresholds were subsequently interpreted within their corresponding crop management periods. Consequently, the figures display monthly calibrated thresholds, while the legends identify the associated crop management periods.

The threshold selection procedure is described in the previous version. Specifically, candidate SESR thresholds were evaluated using an empirical utility-maximisation approach (lines 335–340), in which precision, recall, and Fβ were computed across the tested threshold range. For each crop management period and soil texture class, the threshold that maximised the Fβ score was selected as the optimal calibrated SESR threshold. Subsequently, these optimal calibrated thresholds were subjected to uncertainty and performance evaluation using bootstrap-derived CI95%, FAR, and ROC-AUC (lines 355–366). Finally, only those optimal calibrated thresholds that simultaneously satisfied the predefined performance and uncertainty criteria (lines 367–369) were retained as skilful calibrated impact-informed SESR thresholds. If these criteria are not satisfied, the threshold remains the optimal threshold identified through the Fβ maximisation procedure but is simply not retained as a skilful calibrated threshold. In the revised version, we propose introducing a summary of this sequence at the beginning of Section 3.5 to make the progression from threshold optimisation to skilful threshold identification more explicit.

11. In section 3.3, Is this the criteria (i.e., precision ≥ 0.70, FAR ≤ 0.30, ROC-AUC ≥ 0.60, and CI95% ≤ 0.40) used in selecting the best SESR Threshold? It was not described in the methodology. Please clarify that on top of the method presented of establishing the SESR threshold by maximizing , you have these additional criteria. It is confusing also that several optimal SESR threholds was generated as shown in your Fig. S3 and in this section you retain a certain SESR threshold based on these additional criteria. Clarify and make the method of selecting the optimal SESR threshold coherent with the results presented. Please improve the methodology part to better understand the results being presented.

We thank the reviewer for this observation. The criteria used to identify skilful calibrated impact-informed SESR thresholds (precision ≥ 0.70, FAR ≤ 0.30, ROC-AUC ≥ 0.60, and CI95% ≤ 0.40) were described in the methodology section (lines 367–370 of the previous version). To do so, for each soil texture–month combination, an empirical utility-maximisation approach was used to evaluate a complete range of candidate SESR thresholds. The threshold that maximised the weighted Fβ score was identified as the single optimal calibrated threshold, as illustrated by the solid vertical line in Figure 6. Therefore, the optimisation procedure does not produce multiple optimal thresholds for the same soil texture–month combination; rather, it identifies a single optimal threshold.

The multiple t^*values in Figure S3 arise from the bootstrap procedure. These multiple t^*values are not alternative candidate thresholds from which one is subsequently selected. Instead, they are repeated estimates of the same optimal threshold obtained from different resampled datasets and constitute the distribution presented in Figure S3, from which the 95% confidence interval (CI95) is subsequently derived. Finally, the criteria of precision, FAR, ROC-AUC, and CI95 are not used to select another threshold. Instead, they are used to assess whether the optimal threshold obtained for that soil texture–month combination qualifies as a skilful calibrated threshold. As we mentioned in the previous answer, if these criteria are not satisfied, the threshold remains the optimal threshold identified through the Fβ maximisation procedure but is simply not retained as a skilful calibrated threshold.

In the revised version, we propose adding a short introductory sentence at the beginning of Section 3.5 and using the term “skilful calibrated thresholds” consistently to guide the reader more clearly through the distinction between optimal calibrated thresholds and the subset retained for validation.

12. In section 3.4, It is noteworthy to include the supplemental figures 4 and 5 as main results and not as a supplemental material.

We thank the reviewer for this suggestion. We agree that Figures S4 and S5 provide the quantitative evidence used to compare the performance of literature-based and calibrated thresholds against the two agricultural impact benchmarks. In the revised version we propose moving Figures S4 and S5 from the Supplementary Material to the main manuscript.

Minor comments and comments across the manuscript

- Line 15: Please check if the term is correct. I believed it should be 'soil moisture depletion'.

We thank the reviewer suggestion. The term has been corrected accordingly.

- In line line 32: how do you measure skillful? maybe you can insert a numerical value to quantify this skill of the calibrated threshold.

We thank the reviewer suggestion. We agree that the term "skilful" should be defined when first introduced. In the revised manuscript, we will clarify in line 32 that skilful calibrated thresholds are those satisfying the predefined performance and uncertainty criteria. However, the corresponding methodological definition is provided in Lines 367–372. Define skilful thresholds (precision $\geq$ 0.70, FAR $\leq$ 0.30, ROC-AUC $\geq$ 0.60, and CI95% $\leq$ 0.40).

- In line 33: are you referring to the results of the literature-based approach? if yes, then you can further improve the wordings to specifically refer with results from the literature-based approach. From this text, I am understanding it as a report coming from the literatures which I guess is misleading.

We thank the reviewer suggestion. This will be corrected.

- In line 43: I suggest to include a discussion relating these SDGs with the claim why drought impedes their achievement to expound on the relevance of the study.

We thank the reviewer for this suggestion. Rather than expanding the discussion, we will revise the corresponding text to ensure that the relevance of the study is stated more clearly while maintaining the scientific focus and avoiding unnecessary additional discussion.

- In line 98: Crop management period seems okay in describing the temporal growth stages of crop in terms of the management/cultural practice being implemented (for example planting, fertilization, weeding, harvesting), but maybe give more details on its definition in the context of this study for clarity. I thought at first, it refers to the crop phenological stages or crop growth phases/periods but was later cleared somewhere in the methodology about the term. Maybe as early in these section, you can emphasize that you refer to the crop management period as pre-sowing, sowing, growing, harvesting.

We thank the reviewer for this suggestion. We agree that the term “crop management period” should be defined when first introduced. In the revised manuscript, we will clarify that, in the context of this study, crop management periods refer to the pre-sowing period (April), sowing period (May), growing period (June–July), and harvesting period (August) of the regional cropping cycle.

- In line 102-108 for this objective, with the rigorous literature presented indicating the relevance of SESR in detecting AFDs, what is the need for this analysis?

We thank the reviewer for this observation. We agree that the relevance of SESR for flash drought detection has been widely demonstrated in the literature and that the purpose of Objective 1 was not to re-evaluate SESR as an indicator. Rather, this objective was included because the study focuses on a dryland rainfed environment where high evaporative demand and chronically low soil moisture make evaporative stress part of the normal background conditions. Under such circumstances, literature-based SESR thresholds may not consistently distinguish anomalous evaporative stress from background conditions or correspond to the same levels at which agricultural impacts occur. Consequently, under local dryland conditions, it remains unclear whether SESR exhibits distinguishable responses between years with and without drought-induced crop losses across soil textures and crop management periods. Objective 1 –that is now referred as to "sub-objective" 1–was therefore designed as a diagnostic assessment prior to calibration because impact-informed threshold derivation would lack a sufficient basis if SESR did not exhibit distinguishable responses between years with and without drought-induced crop losses under local conditions. To make this rationale explicit, we suggest adding the following sentence to the Introduction, specifically in lines 123-126: "Moreover, it cannot be assumed that SESR responses are sufficiently distinct to differentiate conditions associated with drought-induced crop losses across soil textures and crop management periods." And then, revise the wording of sub-objective 1 to better reflect this purpose, shifting the focus from evaluating the general agricultural relevance of SESR to assessing whether SESR responses differed between years with and without drought-induced crop losses across crop management periods and soil textures.

- In line 125: typographical error

Thank you, this will be corrected.

- In line 153: I suggest to change the subsection title into 'Data Sources and Preprocessing' to make a distinction and avoid the redundancy of 'Data' as it was already use as section title in Section 2. Data and Methods

We thank the reviewer for this suggestion. In the revised manuscript, the subsection title will be changed to "Data Sources and Preprocessing" to avoid repeating the term "Data", which is already used as the title of Section 2.

- In line 156: please specify what particular qualitative information you are referring to that you extract from news reports.

We thank the reviewer for this suggestion. The qualitative information extracted from newspaper articles and reports includes the date of the reported event, a brief description of the reported impacts, the type of impacts (e.g. agricultural, economic, or population affected), and the corresponding reporting sources. This information was already included in the benchmark construction section but will be moved to Section 2.6 in the revised manuscript to clarify the qualitative information extracted from the reports.

- In line 163 Are you referring to the MODIS MYD16A3GF? please include also the complete specification like temporal resolution.

We thank the reviewer for this suggestion. We confirm that this section refers to the MODIS MYD16A2GF Version 6.1 product rather than the annual MYD16A3GF product. MYD16A2GF provides gap-filled 8-day evapotranspiration (ET) and potential evapotranspiration (PET) estimates at a spatial resolution of 500 × 500 m from 2001 onwards and was used to derive SESR in this study. The revised manuscript will explicitly include the correct product name together with its temporal and spatial resolution.

- In line 163: 500m is a length and is misleading when use to specify spatial resolution. Studies used L x W (500m x 500m) to specify the spatial resolution.

We thank the reviewer for this suggestion. In the revised version we will include this correction.

- In line 164: Please verify if the dataset from MODIS MYD16A3GF is actual evaporation and potential evaporation and not actual evapotranspiration and potential evapotranspiration.

We thank the reviewer for this observation. We verified the specification of the MODIS MYD16A2GF Version 6.1 product and confirm that it provides evapotranspiration (ET) and potential evapotranspiration (PET). The terminology in the previous version was therefore incorrect and will be corrected throughout the revised manuscript.

- In line 166-169: is it possible to derive a flow chart for this process to have better visualization of the process supported by a description for clarity and simplicity.

We thank the reviewer for this suggestion. We agree that the description in lines 166–169 can be simplified to improve readability. Rather than introducing a flowchart, the revised manuscript will streamline this paragraph by removing unnecessary procedural detail and focusing on the main processing steps, while retaining the relevant methodological references.

- In line 171: is the timestep yearly, since the temporal resolution of the MODIS MYD16A3GF is annual?

We thank the reviewer for this suggestion. In the revised version we will include this correction. Please refer to our response to Comment #5, where this issue is addressed.

- In line 179: this unit is misleading when specifying a spatial resolution.

We thank the reviewer for this suggestion. In the revised manuscript, this will be corrected to indicate that the Vegetation Health Index (VHI) dataset has a spatial resolution of 1 km × 1 km.

- In line 180: what is also its temporal resolution?

We thank the reviewer for this suggestion. The VHI has a decadal (10-day) temporal resolution and has been available since 1984 (FAO, 2014). This information will be added to the revised manuscript.

- In line 211: what do you mean by this sampling interval? It is not clear to me how the sampling method is done. Does it mean you selected a 3km x3km grid for each soil texture? the 388 time series is also not clear which time scale it is referring to? I am confuse with the data being processed because the MODIS MYD16A3GF has annual temporal resolution and here you are mentioning an 8-day SESR values? In addition, why do you need to do sampling? is it not possible to process all the pixels covered by each soil texture? was the data too large that takes to long to perform the computations?

We thank the reviewer for this suggestion. Please refer to our response to Comment #5, where this issue is addressed.

- In line 220-221: what is the relevance on why you consider the criteria of having VHI below 0.35 at the end of cropping season as indication of drought years with crop losses? why not use the VHI on the first month of the growing season, or the VHI at the 3 month of the growing season which coincides with the flowering and yield formation stage for maize and the most sensitive stages to drought which could cause yield reduction or crop damage?

We thank the reviewer for this suggestion. Please refer to our response to Comment #6, where this issue is addressed.

- In line 257-265: is this the literature-based AFD detection? If so, I think the title needs to add a term like 'literature-based AFD detection' just to showcase the method. It is also better if a more detailed description on how this approach (literature-based) detect an AFD for a reader who is not familiar with the method. At first reading, I was not able to understand how the method works. A separate section to discuss the proposed enhancement to the existing method is suggested. In this way, as a reader I can separately compare the two approach.

We thank the reviewer for this suggestion. Please refer to our response to Comment #7, where this issue is addressed.

- In line 276: please check the subscript for consistency.

We thank the reviewer for this suggestion. We will correct the subscript of $SESR$ in Equation (3) from s to k so that it is consistent with the notation used throughout the manuscript and the accompanying variable definitions.

- In lines 284-294: this is a results presented in methodology which is unusual.

We thank the reviewer for this observation. In the revised version, we propose relocating this example to a new subsection entitled "Application of the impact-informed calibrated SESR thresholds to detect agricultural flash drought", where it will be presented after the derivation and identification of skilful calibrated thresholds. This reorganisation separates the description of the detection methodology from its application and interpretation.

- In line 235: I suggest a flow diagram/chart for better visualization of the method to reduce the text intensive description.

We thank the reviewer for this suggestion. Rather than introducing a flow diagram, in the revised version we propose using a more descriptive section title and adding a brief introductory paragraph to summarise the sequence of threshold derivation, performance evaluation, uncertainty estimation, and skilful threshold identification. We believe this revision improves readability while retaining the methodological detail required for reproducibility.

- In line 330: Please clarify the MODIS datasets being use because in the section for data, it enumerated the MODIS MYD16A3GF as the data being processed which has annual temporal resolution and not an 8-day resolution.

We thank the reviewer for this suggestion. Please refer to our response to Comment #5, where this issue is addressed.

- In lines 343-345: what are the threshold for this metric to indicate the SESR threshold is performing well?

We thank the reviewer for this observation. We acknowledge that the precision threshold of 0.70 is not a universal performance standard. Rather, it was adopted as a pragmatic decision criterion to identify thresholds that correctly classified the majority of drought-impact years while limiting false alarms. Because no single metric fully characterises threshold performance, this criterion was interpreted jointly with FAR, ROC-AUC, and uncertainty measures. Thresholds were therefore considered skilful only when they simultaneously satisfied all performance criteria rather than based on precision alone.

Precision and recall range from 0 to 1. Precision and recall were first calculated for each candidate SESR threshold and then combined through the Fβ score, which was used as the optimisation criterion. Subsequently, a precision value of at least 0.70 was adopted as one of the criteria for identifying skilful calibrated thresholds. A precision value of 0.70 implies that at least 70% of detected events coincide with years exhibiting drought-induced crop losses, corresponding to a false alarm rate of no more than 30%.

From an operational perspective, the meaning of higher precision is explained in the discussion section: Operational relevance and limitations of calibrated SESR thresholds for flash drought monitoring and management (Lines: 619–672). To improve clarity, we will revise the text in the methodology section to better explain the role of these metrics within the threshold optimisation and selection procedure.

- In line 374: how do you evaluate the performance of the calibrated threshold and the performance of the literature-based approach? what are the metrics you measured to determine whether an improvement is achieved by the calibrated SESR Threshold? what does it mean by V1 =1 or V1 = 0, does it mean a hit or miss? For example, for year 2014, if it is V1=1 then it is a hit meaning it predicted or detected the occurrence of AFD?

We thank the reviewer for this suggestion. Please refer to our response to Comment #8, where this issue is addressed.

- what is crop loss condition = 1 means?

We thank the reviewer for this suggestion. Please refer to our response to Comment #8, where this issue is addressed. We will revise the text to clarify.

- Line 393 : For the results, since literature-based approach have already used the SESR as the parameter in detecting AFD, why do you need to correlate SESR with VHI. A need to expound the relevance of this results in connection to the overall aim of the study.

We thank the reviewer for this suggestion. Please refer to our response to Comment #9, where this issue is addressed.

- Lines 396-399: which correlation between SESR and VHI mostly contributed to the overall correlation when you consider the different soil texture. Does clay soil have leads to a better correlation between SESR vs VHI?

We thank the reviewer for this observation. Figure 4 includes observations from all soil texture classes, which are distinguished by colour. However, the regression equations and correlation coefficients reported in the figure were calculated using the aggregated dataset rather than separately for each soil texture class. The subsequent analyses presented in Figure 5 and Figure S2 indicate that SESR responses were more clearly differentiated between years with and without drought-induced crop losses in clay-rich soils than in sandy soils. We therefore propose moving Figure 4 to the Supplementary Material and retaining Figure 5 as the primary figure, as it more directly supports the objective of assessing whether SESR distinguishes between years with and without drought-induced crop losses across soil textures.

- In line 345: Performance was presented, however, it is also important to showcase the calibrated threshold as it was presented as a supplemental material Fig. S3. A section should be allotted for the presentation of the variation of the calibrated optimal thresholds prior to the to section 3.2. In addition, the context where SESR threshold being calibrated for the different 'crop management period' was not implemented, instead the results showed a calibrated SESR thresholds at monthly basis and not on the time frame for each crop management period.

We thank the reviewer for this suggestion. Please refer to our response to Comment #10, where this issue is addressed.

- Lines 446-448: check the grammar for readability.

We thank the reviewer for this suggestion. We will revise the text.

- In line 450: The calibrated SESR threshold was not clearly describe how it was selected among the incremental runs of maximizing Fb. It was also confusing when I relate it in section 3.3 where it presents plots of optimal SESR thresholds. There were several optimal SESR derived but how did you select the best SESR presented in Fig. 6 among these SESR?

We thank the reviewer for this suggestion. Please refer to our response to Comment #10, where this issue is addressed.

- Line 486: Is this the criteria for selecting the best SESR Threshold? Please clarify that on top of the method presented of establishing the SESR threshold by maximizing Fb, you have this additional criteria. It is confusing also that several optimal SESR threholds was generated as shown in your Fig. S3 and in this section you retain a certain SESR threshold based on this additional criteria. Clarify and make the method of selecting the optimal SESR threshold coherent with the results presented. Please improve the methodology part to better understand the results being presented.

We thank the reviewer for this suggestion. Please refer to our response to Comment #11, where this issue is addressed.

- Line 491: what does it mean by skillful threshold?

We thank the reviewer for this suggestion. Please refer to our response to Comment #11, where this issue is addressed.

- Lines 513-515: It is noteworthy to include this as main results and not as a supplemental material.

We thank the reviewer for this suggestion. Please refer to our response to Comment #12, where this issue is addressed.

- Line 579: please check the term. is it referring to low water holding capacity?

We thank the reviewer for this observation. The term “low water capacity” has been revised to “low water-holding capacity”.

- Add the perspective of Figure about window for early action. https://www.preventionweb.net/media/93703/download?sta

We thank the reviewer for this suggestion. The discussion already addresses the relationship between flash drought detection, the timing of agricultural decisions, and the trade-offs between early action and false alarms. In the revised manuscript, we will make this perspective more explicit by linking the discussion to the window for early action framework.