RESPONSE TO REVIEWER #1 (Samuel Jonson Sutanto)

We would like to express our gratitude to Samuel for taking the time to review our paper and for providing valuable feedback and suggestions. We appreciate the thoroughness of your review, which has significantly contributed to improving the quality of the manuscript.

Below is a point-by-point response to all comments. Original comments are in black, whereas the authors' responses are in blue.

Title: Indicator-to-impact links to help improve agricultural drought preparedness in Thailand

Authors: Tanguy et al.

Recommendation: minor revision

Summary

This paper correlates meteorological drought indices, represented by SPI and SPEI, and vegetation indices such as VCI, TCI, and VHI with forest growth and crop yield impacts. Two approaches were used in the analysis, which are the Pearson correlation and the Random Forest machine learning model. The authors found that the strength of correlations depends on land use, season, region, and drought duration. Crops are strongly impacted by drought in both wet and dry seasons. The impact of droughts, however, is less apparent for forest growth. The use of the Random Forest technique allows a more in-depth analysis of the importance of different drought and vegetation indicators. The authors also highlighted that the knowledge of linking specific indicators to the drought's impact on crops will help to improve the DMEWS and perform mitigation actions.

Assessment

This paper analyzes the use of different drought and vegetation indicators to link these indices with the impact of drought on crop yields and forest growth. The manuscript is interesting and well written. I have a few minor comments below and two general comments, but only for clarification. I believe this work is well suited for NHESS.

Thank you for the encouraging and positive feedback.

General Comments

I have two general comments regarding the manuscript but all of them are only for clarification and improvement of the manuscript.

1. I am wondering why the authors used the Random Forest (RF) approach only to find the importance of all indicators on crop yields and forest growth. RF can also be used to predict the crop yields by: 1) training the predictor variables, here are e.g., drought and vegetation indices, and the response variables (e.g., crop yields), resulting in crop yield impact model; 2) using the developed model to predict the impact of drought on crop yield by leaving out the predicted year from the training period. Maybe it is interesting to do this since the authors already have the script to develop the RF model and mentioned this in Figure 2. The authors can train the RF model again without the predicted year and in the end forecast the yields and validate the result with the observed crop yield data. Otherwise, it is worth to discuss the use of machine learning to predict crop yield and not only to find the importance of predictor variables.

Thank you for your comment. We agree that RFs can be used to predict crop yield and we have indeed used it in that way, as mentioned in line 221, where we state that "Regional Random Forest (RF) models were used to predict agricultural impacts (crop yield)." However, we appreciate the suggestion to explore this application of RFs more explicitly in our manuscript. To make it clearer that the RFs were used to predict crop yield, although this was not the main focus of our study, we will add the following sentence to the end of section 2.3.2:

"*While RFs were built to predict crop yields, the main focus of our study is their use to study feature importance to identify monitoring priorities for different regions and crops*."

Regarding the suggestion to use leave-one-out cross-validation, we used 5-fold cross-validation instead, as we explained in lines 243-245. Our choice was based on computational efficiency, as 5-fold cross-validation is faster than leave-one-out cross-validation.

We also agree that we could have emphasised more the potential of RFs to predict drought impacts (crop yields in our case). To address this, we will rephrase the first paragraph of the "Future Work" section (line 511-514) as follows:

"*In this study, we used RF models primarily to analyse the relationships between drought indicators and impacts, and to identify the relative importance and timing of relevant indicators. While the main focus of our analysis was on feature importance, our analysis also demonstrated the potential of RFs to simulate unseen data, which suggests they could be used for impact prediction. With further work, such as* *addressing the limitations discussed above, these models could be used for DMEWSs, support and compensation schemes, long-term planning, etc."*

2. I have difficulty to understand figures 8, 9, and 10. I read the caption over and over again but still cannot interpret the figures. Is there any other way to present your results in a simple manner, so thus the readers can understand the results? For example, it is not clear to me why some lines are thick, and some are thin. Also, why VCI N has 6 thin lines and VCI E has only 3 thick lines? How to indicate 24 months accumulation periods in the results? Maybe modify the Y-axis?

Thank you for your valuable feedback regarding the difficulty in interpreting figures 8, 9, and 10. We acknowledge the deficiencies in these plots and have carefully considered the best way to present our results. We have developed two alternatives, one that is a slightly modified version of the existing plots and a second that is an alternative version of the heatmaps. We will present both here, but we plan to use the modified version of the existing plots for the revised version. We will also provide a more detailed explanation in the caption to improve the interpretation of the figures.

Regarding the number of lines in the plots, we apologize for the lack of clarity in our original caption. Figure 8 shows the feature importance for all variables used in all models within each region. The number of variables per model and the number of models vary across regions and crops, resulting in different numbers of lines for each subplot. Specifically, the number of crops modelled per region varies, and RFs were not able to model some of the crops in some of the regions. For example, in the N region, we modelled 5 crops, but in the NE region, we were only able to model one crop (cassava). We will revise the caption to provide a clearer explanation.

The thickness of the lines does not have any meaning attached to it and only reflects the number of variables in that subplot. There are more variables used in all the models in the N region than in the other regions, resulting in thinner lines in the N subplot. An option would have been to have different sized subplots to have lines with equal thickness. However, this solution would have been much more complicated technically, as we have used python module *seaborn*'s 'heatmap' functionality, which doesn't have that option built in.

We also understand that it was not clear why VCI (for example) has multiple lines in the plots. We separated the VCI for each individual month separately (see line 225), and sometimes VCI for the same month is repeated because it was used in different models for different crops. We will add a clearer explanation in the revised caption.

Finally, we acknowledge that the accumulation period of 12, 18, and 24 months could not be differentiated in the original plots. In the revised version, we have added the full variable name explicitly, so the accumulation period can now be identified.

Thank you again for your feedback, which has helped us improve the clarity and interpretation of our results.

Alternative Figure 8:

Option 1: We have revised Figure 8 by explicitly labelling each indicator to improve its readability (y-axis). This is the only difference between the revised and original versions. We acknowledge that the labels for Region N may appear small, but we hope that readers can still discern them by zooming in. We believe this is preferable to omitting the labels altogether, as was done in the original version. Additionally, we have updated the caption with more detailed explanations to clarify the information presented in the figure.

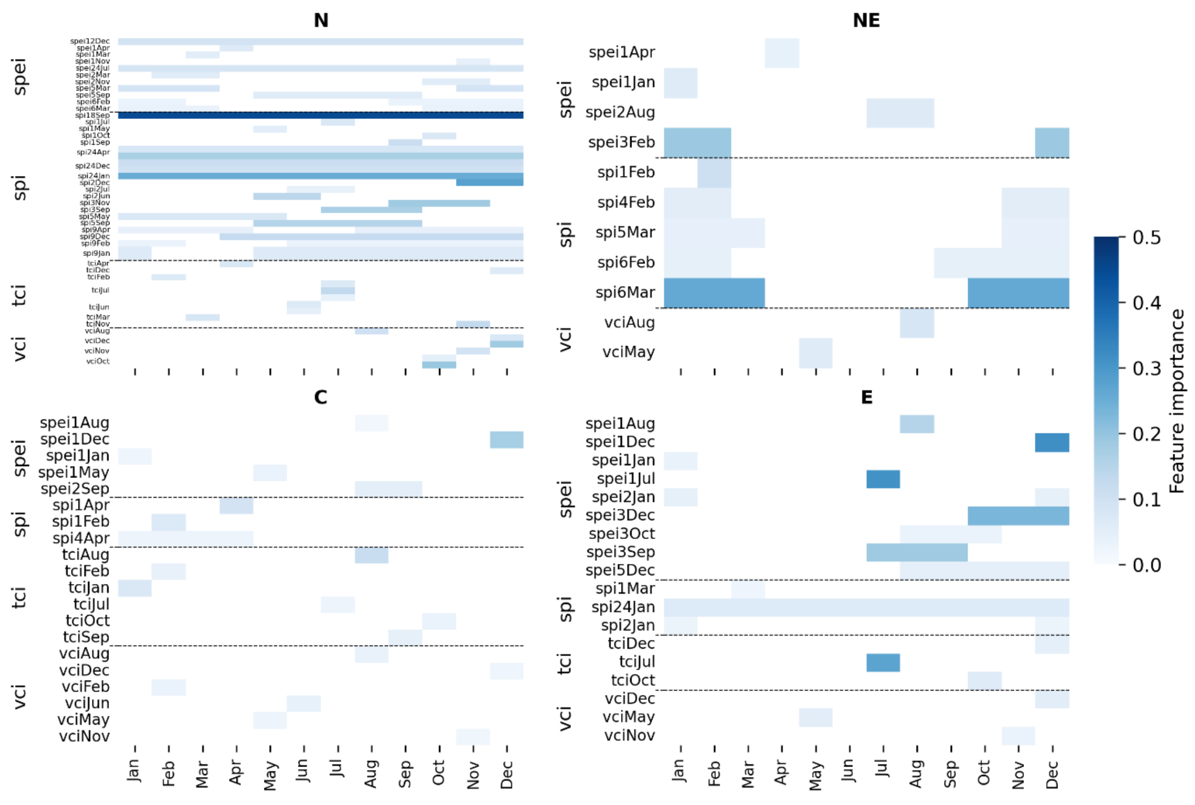

Figure 8: Heatmap displaying the relative feature importance (impurity decrease) of each indicator used in the random forest models (for all crops) for each region. Each row corresponds to a different indicator, with the y-axis representing the indicator and the length of the bar representing the accumulation period. The x-axis indicates the time of year (month) when the indicator is most relevant for predicting crop yield. For instance, spi6Mar in the NE region represents SPI with a 6-month accumulation period for March, and the bar covers October to March. The bars are shaded darker for indicators that are more important in the models. Unlike Fig. 9 and 10, which show only one crop per subplot, this figure includes all crops that can be modelled in each region. Region N has five models (cassava, corn S1, corn S2, mixed corn, and paddy rice models), while Region NE, Region C, and Region E have one, one, and two models respectively. The number of lines (i.e., indicators) in each subplot is a consequence of the number of models in each region and the number of variables in each model. The thickness of the lines is a result of the number of indicators displayed and has no meaning attached. Finally, note that different crop models within a region can use the same indicators, leading to multiple lines for some indicators (e.g., vciDec in Region N).

Option 2: We attempted an alternative method of displaying the information in our plots. In these visualisations, we included all indicators, including those that were not used to construct the random forest models (which are greyed out). This approach ensures that each subplot has the same number of indicators and avoids the issue of varying line thicknesses. However, this method omits an important dimension: the period covered by the accumulation period. We rely on this information extensively in our text to explain differences between regions and crops. Consequently, we have decided to retain option 1 for our plots.

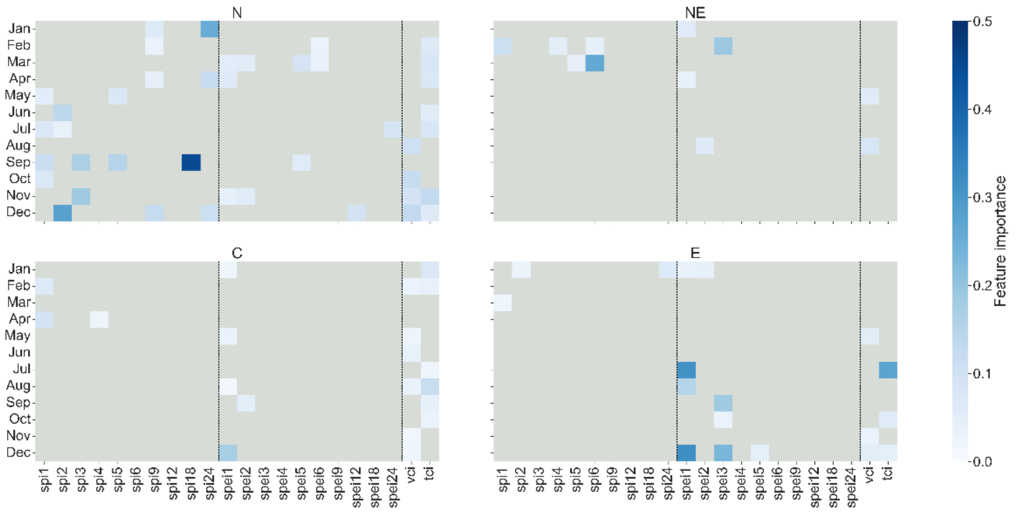

Figure 8: Heatmap depicting the importance of each feature (indicator) for all models in each region. The x-axis shows the indicators, with the accumulation period included where relevant, and the y-axis shows the feature importance for each month of the year. Greyed areas represent indicators that were not used to build any of the models in that region. Unlike Fig. 9 and 10, where only one crop is shown per subplot, each subplot in this figure includes all crops that could be modelled in that region. Region N has five models (cassava, corn S1, corn S2, mixed corn, and paddy rice models), while Region NE, Region C, and Region E have one, one, and two models, respectively.

Alternative Figure 9:

Option 1: Similar to the new Figure 8 (option 1), the only difference between this figure with the original version is the explicit labelling of each indicator, and improved caption.

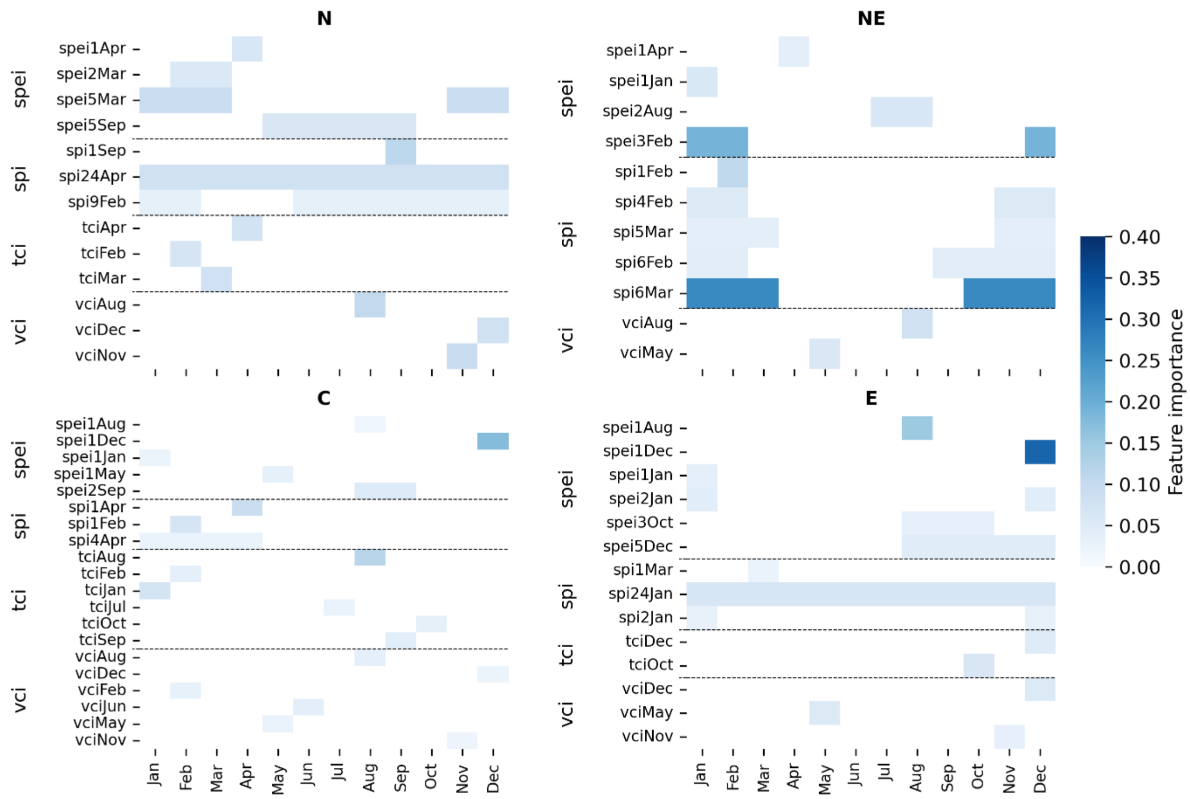

Figure 9: Heatmap displaying the relative feature importance (impurity decrease) of each indicator used in the random forest models for Cassava for each region. Each row corresponds to a different indicator, with the y-axis representing the indicator and the length of the bar representing the accumulation period. The x-axis indicates the time of year (month) when the indicator is most relevant for predicting crop yield. For instance, spi6Mar in region NE represents SPI with a 6-month accumulation period for March, and the bar covers October to March. The bars are shaded darker for indicators that are more important in the models. Unlike Figure 8, each subplot here shows only cassava models for each region. However, the number of indicators can still differ between models due to the feature selection process that eliminates highly correlated indicators, which may vary between regions. The number of lines (i.e., indicators) in each subplot reflects the number of variables in each model, and the thickness of the lines is a result of the number of indicators displayed and has no meaning attached.

Option 2: Similar to the alternative option 2 for Figure 8, this heatmap has the advantage of displaying an equal number of indicators in each subplot, which can facilitate the comparison of feature importance between regions. However, this alternative visualisation does not include information on the accumulated period of the indicators, which is a relevant aspect for our analysis and discussion. Therefore, we chose to use the improved version of our original heatmap (option 1) in our study, as it provides a more comprehensive representation of the feature importance for each indicator, including its temporal relevance for predicting cassava yield.

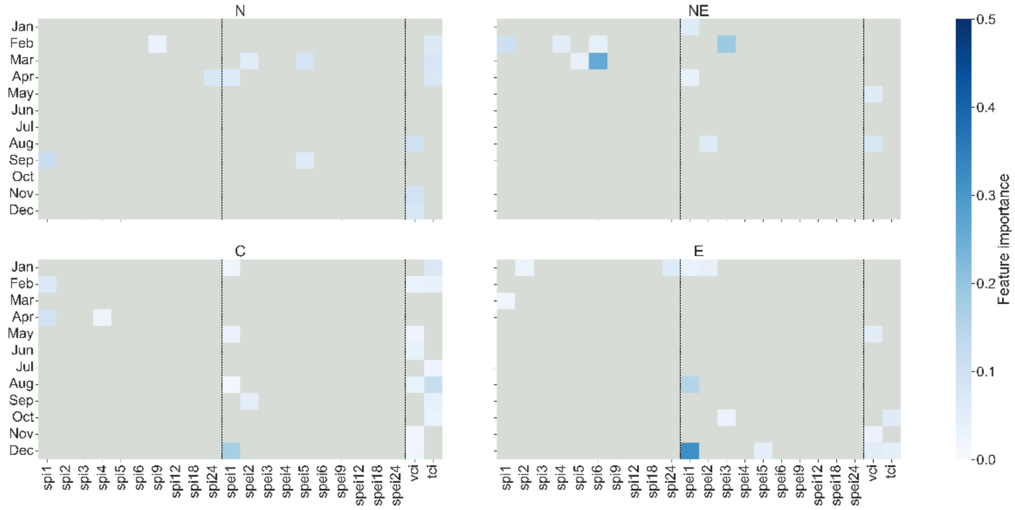

Figure 9: Heatmap depicting the importance of each feature (indicator) for cassava models in each region. The x-axis shows the indicators, with the accumulation period included where relevant, and the y-axis shows the feature importance for each month of the year. Greyed areas represent indicators that were not used to build the cassava model in that region. Note that despite modelling the same crop in each region, the number of indicators can differ between models due to the feature selection process that eliminates highly correlated indicators, which may vary between regions.

Alternative Figure 10:

Option 1: Similar to the new Figure 8 (option 1), the only difference between this figure with the original version is the explicit labelling of each indicator, and improved caption.

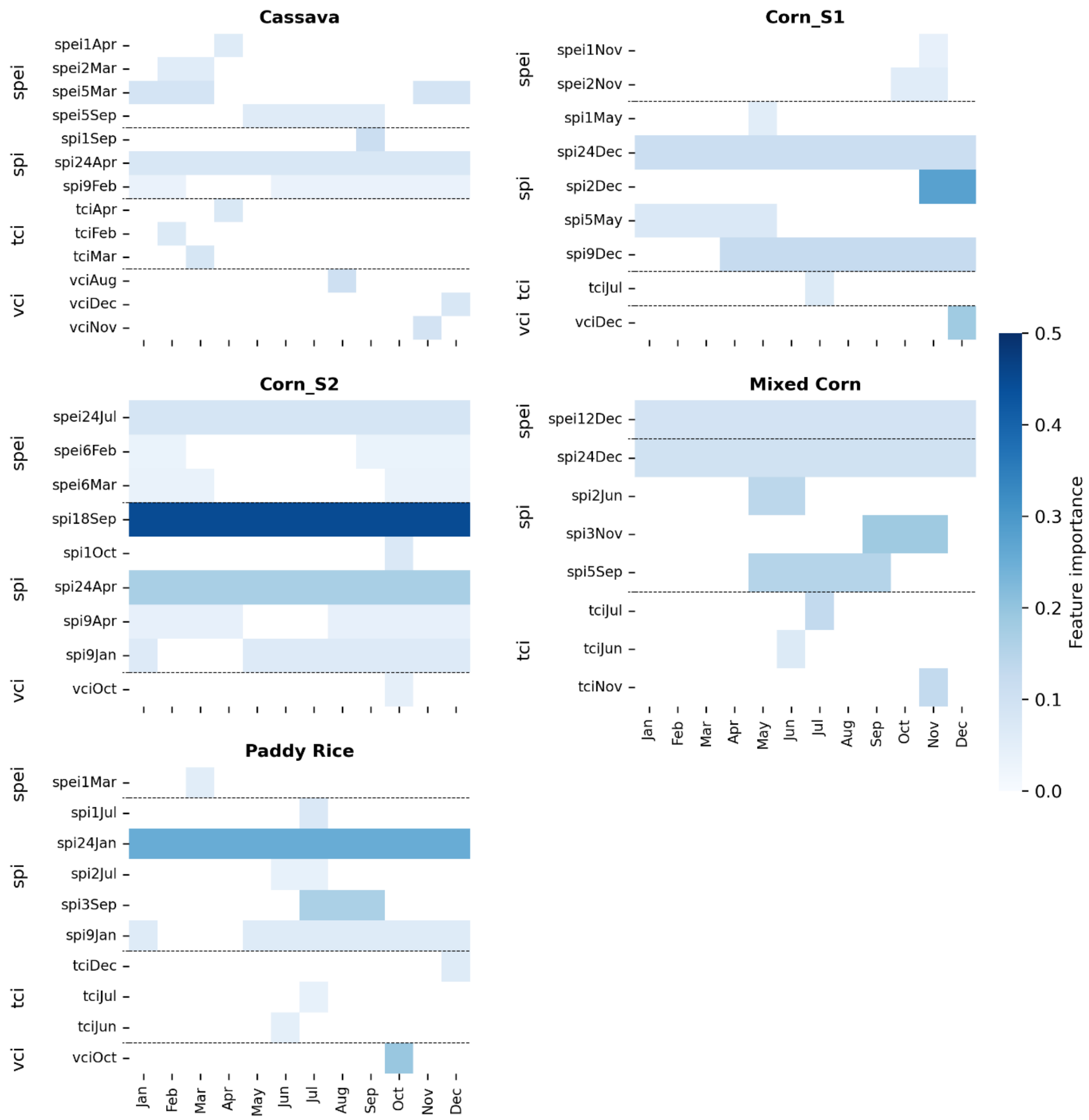

Figure 10: Heatmap displaying the relative feature importance (impurity decrease) of each indicator used in the random forest models for five different crops in region N. Each row corresponds to a different indicator, with the y-axis representing the indicator and the length of the bar representing the accumulation period. The x-axis indicates the time of year (month) when the indicator is most relevant for predicting crop yield. For instance, spi5Sep for Mixed Corn represents SPI with a 5-month accumulation period for September, and the bar covers May to September. The bars are shaded darker for indicators that are more important in the models. Unlike Figure 8, each subplot here only shows the model for a single crop in region N. The number of lines (i.e., indicators) in each subplot reflects the number of variables in each model, and the thickness of the lines is a result of the number of indicators displayed and has no meaning attached.

Option 2: Same comment as Fig. 9, option 2.

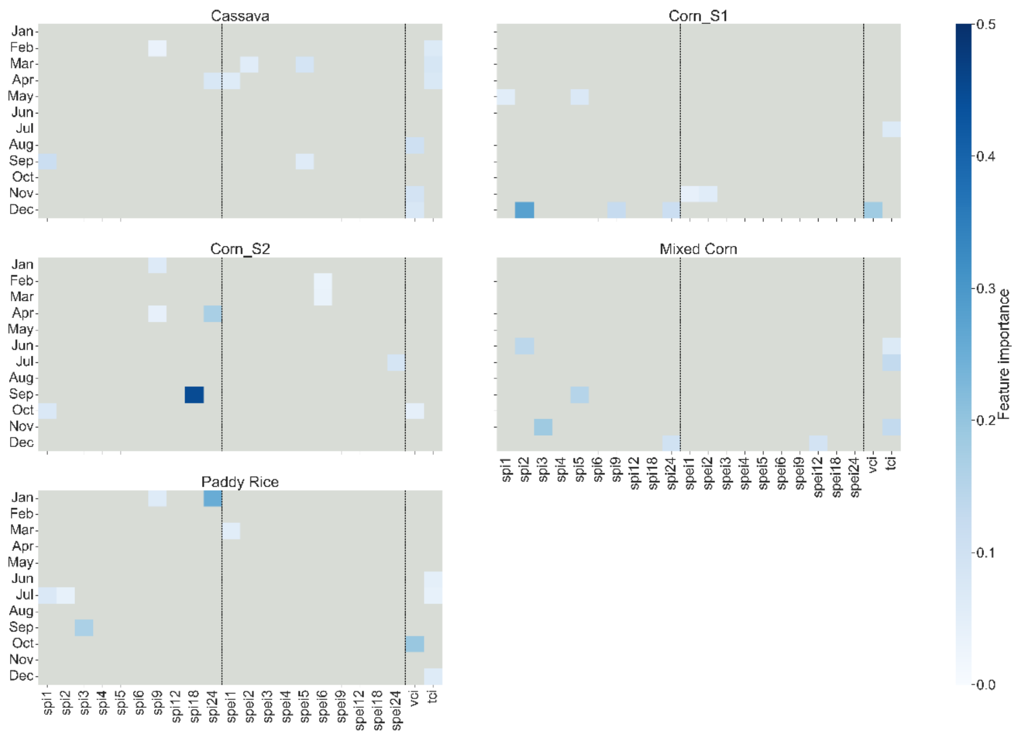

Figure 10: Heatmap depicting the importance of each feature (indicator) for RF models for five different crops in region N. The x-axis shows the indicators, with the accumulation period included where relevant, and the y-axis shows the feature importance for each month of the year. Greyed areas represent indicators that were not used to build the cassava model in that region.

Line by line comments

L refers to line and P refers to page.

P1L19: Maybe re-write "…it provides stakeholders…" as "…provided to stakeholders…"?

Thank you. This will be corrected in the revised version.

P2L33: The authors may add a study on extreme high and low flow events in Southeast Asia including Thailand due to climate change (Hariadi et al., 2023).

Thank you for your suggestion. We agree that this study would be valuable in setting the context for our research by highlighting the expected impact of climate change on high and low flows in the region. We will incorporate the reference you provided into our revised manuscript.

P2L46: Full stop after the ICID reference.

Thank you, this will be corrected in the revised manuscript.

P2L53: “has” -> “have”

Thank you for spotting this mistake, we will correct it in the revised manuscript.

P3L75: Double reference from Stahl et al., 2016.

Thank you for pointing this out. We apologise for the confusion. The reference in question actually comprises two different sources: Bachmair et al., 2016 and Stahl et al., 2016. It appears that our citation management software incorrectly displayed the first one as Bachmair, Stahl et al., 2016. We will rectify this issue in the revised manuscript to ensure accurate and clear referencing.

Bachmair, S., Stahl, K., Collins, K., Hannaford, J., Acreman, M., Svoboda, M., Knutson, C., Smith, K. H., Wall, N., Fuchs, B., Crossman, N. D., & Overton, I. C. (2016). Drought indicators revisited: the need for a wider consideration of environment and society. WIREs Water, 3(4), 516-536. https://doi.org/https://doi.org/10.1002/wat2.1154

Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acácio, V., Dias, S., Stagge, J. H., Tallaksen, L. M., Kampragou, E., Van Loon, A. F., Barker, L. J., Melsen, L. A., Bifulco, C., Musolino, D., de Carli, A., Massarutto, A., Assimacopoulos, D., & Van Lanen, H. A. J. (2016). Impacts of European drought events: insights from an international database of text-based reports. Nat. Hazards Earth Syst. Sci., 16(3), 801-819. https://doi.org/10.5194/nhess-16-801-2016

P3L76: Better to place the EDII and DIR references here. EDII: Stahl et al., 2016 and DIR: Smith et al., 2014?

Thank you for your suggestion. We agree that providing references to the EDII and DIR here would enhance the clarity of the manuscript. We will make the suggested change and addition in the revised version.

P4L109: I suggest to mention again the gap (instead of “that” gap) here since it is a new paragraph.

Thank you for the suggestion. In the revised manuscript, we will mention explicitly the gap again. We will replace this sentence:

“In this paper, the ambition was to fill that gap, with a focus on agricultural drought impacts at the national scale, across different crops and seasons, comparing the relative utility of traditional statistical methods at high resolution (remote sensing data at provincial scale) vs. lower resolution sectoral specific analyses (applying machine learning approaches to regional/provincial yield data), to inform improved approaches for national DMEW.”

With the following two sentences:

"In this paper, the ambition was to fill the gap in the literature on comprehensive studies investigating the links between drought indicators and impacts at a national scale in Thailand. Specifically, we focused on agricultural drought impacts, considering different crops and seasons, and compared the relative utility of traditional statistical methods at high resolution (remote sensing data at provincial scale) vs. lower resolution sectoral-specific analyses (applying machine learning approaches to regional/provincial yield data), to inform improved approaches for national DMEW."

P5L144: Suggestion to rephrase the sentence: "….of Thailand. Although it has suffered….decades, there has been some…"

Thank you for the helpful suggestion. We agree that the suggested modification enhances the paragraph's readability.

P8: Figure 2. Here the authors clearly indicated that the RF model can be used to predict the crop yield (my general comment 1). This is one thing that I miss from the result.

As highlighted in our response to your general comment 1, we have used RF models primarily to study feature importance to better understand the link between drought indicators and drought impacts. However, we recognise the importance of using these models for impact prediction as well. To empathise this more, we have modified the first paragraph of the "Future Work" section (line 511-514) as follows:

"*In this study, we used RF models primarily to analyse the relationships between drought indicators and impacts, and to identify the relative importance and timing of relevant indicators. While the main focus of our analysis was on feature importance, our analysis also demonstrated the potential of RFs to simulate unseen data, which suggests they could be used for impact prediction. With further work, such as addressing the limitations discussed above, these models could be used for DMEWSs, support and compensation schemes, long-term planning, etc.*"

P9L8: Please elaborate more how the authors did "detrended". The authors only said using a simple linear regression.

Thank you for highlighting the need to elaborate the way detrending was performed in our study. We will clarify on lines 208-210 and 227 that variables were regressed against time using linear regression, and the residuals used to remove linear trends in the data. These are the edits in our manuscript:

Lines 208-210:

Before: "*The monthly crop-masked VCI was detrended before the correlation analysis (using a simple linear regression) in order to remove long-term trends, accounting for increased biomass from developments in agricultural technology and practices.*"

After: "*Monthly crop-masked VCI values were regressed against time using linear regression, and the residuals used in this analysis to remove linear trends, accounting for increased biomass from developments in agricultural technology and practices.*"

Line 227:

Before: "*All input data were first de-trended using a simple linear regression*"

After: "*All input data were first regressed against time, and the residuals used as input to the random forest models to account for linear trends.*"

P10L217: "…both "indices" in our…….

Thank you for the correction, this will be added in the revised manuscript.

P14L291: I am wondering, it is VI or VCI?

Thank you for pointing out this potential confusion. To clarify, VI stands for Vegetation Index (as noted in line 90), while VCI stands for Vegetation Condition Index, which is one of the VIs we used in this study. The statement on line 291, "*For crops, we find high correlations between VI and SPEI of relatively short accumulation period during the dry season*", applies to both VIs considered in this study (VCI shown in the main manuscript, and VHI shown in the supplementary material). However, to avoid further confusion, we will revise the manuscript to only mention VCI in this sentence.

P14L294-296: Maybe elaborate more about the meaning of positive and negative correlations between VCI and meteo indicators. Also, the authors stated that short droughts are beneficial for forest growth. In my opinion, drought is never beneficial for any ecosystem. I suggest to rephrase the word beneficial.

Thank you for bringing up the need to elaborate on the meaning of positive and negative correlations between VCI and meteorological indicators and to rephrase our use of the word 'beneficial'. We agree with your points and will provide further clarification in the revised version. Specifically, we will add the following text:

"*A positive correlation between VCI and the meteorological indicators suggests that a deficit in water availability (as indicated by negative SPI or SPEI) leads to a decline in vegetation growth (reduced VCI). In contrast, a negative correlation suggests that such a deficit leads to an increase in vegetation growth. This second scenario may seem counterintuitive, but it can occur in energy-limited environments where water is not the* *limiting factor. In such cases, short droughts (i.e., periods drier than usual for the time of year) can stimulate increased vegetation growth, as droughts in energy-limited environments are often associated with increased radiation (i.e. energy) due to decreased cloud cover. This is discussed further in section 4.3.*"

P14L297: The authors can consider to re-write "…the accumulation period best correlated is…" as "…the best correlated accumulation period is…"

Thank you for the suggestion. We will change this in the revised version.

P14L298: Here and also in the discussion, the authors conclude that forest is more resistant to short droughts. I believe that this strongly relates to the ability of forest to subtract water from deeper layers, e.g. groundwater. Discuss this.

Thank you for bringing this to our attention. We agree that the deeper root systems of forest trees allow them to extract water from deeper layers of soil, making them more resilient to droughts compared to most crops. However, we have overlooked this point in our manuscript. To address this, we will include this explanation in the discussion section (4.3), along with references such as Breda et al. (2006) and Schenk and Jackson (2002), which support the role of deep roots in conferring resilience to droughts in forests.

Breda, N.; Huc, R.; Granier, A.; Dreyer, E. 2006. Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences. Annals of Forest Science. 63(6): 625-644.

Schenk, H.J.; Jackson, R.B. 2002. The global biogeography of roots. Ecological Monographs. 72(3): 311-328.

P18L343-346: How to see the SPI24 from Figure 9 and to see 11 SPEI, 10 VCI, and 6 TCI from Figure 11? See my general comment 2.

We hope that the revised versions of Fig. 8-10 have addressed this issue.

P22L397: The authors stated that SPI is more important than SPEI. I am wondering whether the low precipitation in the N region has something to do with the result.

Thank you for your comment. This is an interesting point. The low precipitation in region N leads to the Actual Evapotranspiration (AET) to be water limited (i.e. AET < PET) and therefore SPEI could be less closely linked to agrometeorological conditions. However, this is also the case for region W and C, and to lesser extend NE as well, where precipitation is also low, but in these other regions, SPEI's importance is mostly dominant (Fig. 5c). We believe the dominant importance of SPI in region N is linked to the reliance of water storage for irrigation in this region, particularly for Corn_S2 which is planted in the dry season and relies heavily on irrigation (Fig. 10). Therefore, a deficit in rainfall (and consequent depleted storage) will have a strong impact on crop yield. We will add these reflections in the revised manuscript.

P22L410-411: You may discuss the difference in water consumption by each crop.

Thank you for your comment. We appreciate your suggestion to discuss the difference in water consumption by each crop. The irrigation requirement indeed varies greatly between crops. Paddy rice is the most water-intensive crop, with an irrigation requirement of around 520m^3/ton if cultivated during the wet season and 1140m^3/ton in the dry season. Corn, on the other hand, requires irrigation only if cultivated in the dry season, with an irrigation requirement of approximately 850m^3/ton. Finally, cassava is the least water-demanding crop, with an irrigation requirement of around 20m^3/ton in the wet season and 65m^3/ton in the dry season (Gheewala et al., 2014).

We will make sure to include these numbers in our manuscript to provide a more comprehensive understanding of the impact of drought on crop yields. Thank you for this helpful suggestion.

Reference:

Gheewala, S.H.; Silalertruksa, T.; Nilsalab, P.; Mungkung, R.; Perret, S.R.; Chaiyawannakarn, N. Water Footprint and Impact of Water Consumption for Food, Feed, Fuel Crops Production in Thailand. Water 2014, 6, 1698-1718. https://doi.org/10.3390/w6061698

P23L437: Rephrase “though this effect is highly variety specific:

This will be rephrased as: “However, it should be noted that this effect varies significantly depending on the specific crop variety, [...]”

P24L453-455: Make two sentences.

Thank you for the suggestion. We agree that the sentence is too long and we will split it into two sentences as follows:

“*For the crops where it was possible to build a RF model, the analysis of the temporal variation in feature importance and the indicator-to-impact relationships provide insights into critical periods for early warning of impacts and relevant accumulation. Specifically, these are periods of interest when dry conditions could lead to impacts.*”

P24L459: “...seasons, which suggests... -> “...seasons, suggest...”

Thank you, this will be changed in the revised manuscript.

P24L468-470: Explain this already in the beginning, thus the readers will not be confused.

We have mentioned this as a response to your comment on **P14L294-296** earlier. Hopefully this addresses this point too.

P25L498: Here, the authors can link the short drought events with the limitation of using data-driven model, such as machine learning.

Thank you for your comment. To clarify, did you mean 'short drought events' or 'short period of record'? Assuming you meant the latter, we agree that the limitation of using data-driven models such as machine learning is the need for a large amount of data to train the model effectively. In our study, we had a relatively short period of record, which limited the amount of data available for training the models. As a result, the models may not have been able to accurately capture the full range of conditions that could occur in the real world. For example, for species such as longan, which are more susceptible to long drought events, the limited instances of these events in our training data may have affected the model's ability to accurately predict impacts. We will add this discussion to the revised manuscript.

P25L505-506: Rephrase "Though powerful tools to produce predictive models from data"

We will rephrase this sentence as follows:

"*RFs are powerful tools for producing predictive models from data, but they are considered 'black boxes' since they do not explicitly extract the relationships between input features and the predicted outcomes. However, RFs can aid in the interpretation of the model through the analysis of feature importance, which identifies the most influential variables in making predictions.*"

References:

Hariadi et al. (2023). A high-resolution perspective of extreme rainfall and river flow under extreme climate change in Southeast Asia, https://doi.org/10.5194/hess-2023-14.

Smith et al. (2014). Local observers fill in the details on drought impact reporter maps, https://doi.org/10.1175/1520-0477-95.11.1659.