An interesting paper and topic. From an hydrologists perspective there are very limited explanation on the hydro and meteorological input. How good the are and how they are handled. Are they at the same resolutions (0.1 and 0.01 deg) or are they scaled. In case how? They are the most sensitive parameters, but their quality is not discussed. Are the results better where there is a good coverage of observations?

We want to thank the reviewer for reviewing our manuscript, the reviewer's comments have provided us with valuable insights and suggestions that will improve the quality of this manuscript for future readers.

We acknowledge that the discussion would benefit from adding how the quality of the input data may affect the predictive ability of the model presented in our manuscript. Therefore, in the revised version, we have included an additional emphasis on how spatial variability of input data quality may affect our predictions. More specifically, we have relied on the initial evaluation results presented in the literature describing the various input data. For example, when considering the meteorological forcing, we will compare how spatial variations in error of (Muñoz-Sabater et al. 2021 & Singer et al. 2021) may be correlated with the spatial variation of errors in our predictions. In addition, we can make similar comparisons for the landcover types data (Friedl, Mark and Sulla-Menashe, Damien 2019). Please see line 330-337.

Regarding the reviewer's comments on how we treated the spatial resolution differences in the input data, we did attempt to include this information in the original submission. On lines 90-92 we state that: 'In the cases where the data were not available at 0.1° resolution, the nearest-neighbour interpolation scheme from xarray (Hoyer and Hamman, 2017) was used to match the variables to the same spatial resolution.' & in Table 1 we provide information on the original resolution of the input data. In hindsight, and in agreement with the reviewer's comment, we acknowledge that this information is ambiguous and not abundantly clear.

We have provided additional emphasis on this aspect of the methodology in the revised manuscript. We have added additional context on line 84 -95 where we explain how input data at different resolutions were handled to produce the datasets used in this study. We aimed to make it clear that, during the training step all data were provided at a 0.01° resolution. Whereas, during the prediction part of the process we matched all the inputs to the resolution we were predicting at. In other words, when predicting at the 0.01° resolution all input data were also at the 0.01° resolution and when predicting at the 0.001° resolution all input data were also at the 0.001° resolution.

References:

Friedl, Mark and Sulla-Menashe, Damien: MCD12Q1 MODIS/Terra+Aqua Land Cover

Type Yearly L3 Global 500m SIN Grid V006,https://doi.org/10.5067/MODIS/MCD12Q1.006, type: dataset, 2019.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth System Science Data, 13, 4349–4383, 2021.

Yamazaki, D., Ikeshima, Daiki, Sosa, Jeison, Allen, George H., Bates, Paul D., and Pavelsky, Tamlin M.: MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset, Water Resources Research, 55, 5053–5073, 2019.

A better analysis on where (spatially) the results are good and bad would also be useful. Is the model more reliable in Europe than in amazonas for instance? The plots are very coarse and do not give a good view on this…if there are any relations.

In our initial analysis, in which we compared landcover types on the global scale (Figure 6 and Section 4.5), we assumed that comparisons between specific regions would become implicit. For instance, in Section 4.5 we detail how results differ between landcover types, but without making explicit references to specific regions of the world. Furthermore, and in agreement with the reviewer, we intended that Figure 5 provides the reader with comparisons between different regions of the world. We think that by adding additional references to specific regions presented in Figure 5 in Section 4.5, the quality of the manuscript has been improved by providing additional context. Please see the updated Figure 5.

Line 50: What is bad weather… for some rain is bad for others that is good.. Cloudy conditions…

Bad weather in this instance refers to cloud cover that interrupt the transmission and reception of signals between the Earth’s surface and Terra and Modis Sensors. In addition to cloud cover, signal transmission may also be interrupted by pollution in the atmosphere or even technical issues on board the satellites. In the revised manuscript, this sentence, on lines 51-52, reads: “ Revisiting times for Landsat and Sentinel-2 are further prolonged when sensors or retrievals are interrupted by cloud cover, pollution in the atmosphere or even technical issues. “

Line 54-55: Before stating this the argument for why hig spatial and temporal resolution is needed.

We acknowledge that by adding this, the relevance of the manuscript would become clearer to readers. One line of argument that is appropriate here is that effects of changes in vegetation are often felt at fine scales and that coarse resolution data fails to capture these effects. This necessitates the need for high-resolution products that can capture fine-scale vegetation dynamics; thereby providing policy and decision makers with insights, predictions, or forecasts at a relevant scale. We have included this argument in the first line of the introduction, line 21-32.

Line 71-73: Something wrong with this sentence

We agree with the reviewer that this is a confusing sentence. This sentence has been replaced with "This study aims to further our understanding on how well ML methods can be used create vegetation products that are useful for global drought impact applications. We set out to establish whether ML methods are able to alleviate missing data and resolution limitations of remote sensing-based vegetation health products by linking vegetation condition (evi) with meteorological and hydrological data.", line 72-76.

Line 81- 84: This is more or less repeating the previous terxt from ln 73

We agree with the reviewer that this is repetitive. The sentences in lines 81-84 serve to guide the reader through the structure of the method sections by highlighting that each subsection relates to an aim or objective that we stated in the introduction. The lines starting in 73 were first and foremost to specify our aims and objectives; whereas, the lines in 81 are more or less there to make it clear to the reader how our methodology relates directly towards addressing each one of our aims. In our opinion, these lines add clarity even though they are at the cost of being a bit repetitive.

Table 1: What are the resolutin on tp, elv, slope etc.... and how are they aggregated to the same resolution

We believe that the information on the original resolution is already present in Table 1. We refer the reviewer to our previous replies, in which we detail how we aimed to clarify how and when different resolutions were used during training and prediction.

Line 152 -156: Is this better placed in the results section. Is this something that deserves discussion. Internal correlation can lead to exclusion, but aspect is no, and is a logical parameter to influence vegetation at smaller scales 0.1 - 0.01 deg. And was it not scale sensitive?

We acknowledge that aspect may sound like a logical parameter to influence evi values. The random forest algorithm is prone to overfitting, which can lead to poor predictive performance when the algorithm is faced with novel data. To avoid this pitfall, we relied on the correlation analysis presented in Section 2.1.2 and Appendix A1 to identify the variables that are the most appropriate. Given that aspect did not exhibit a strong correlation with evi, we excluded it from further analysis. A weak correlation between aspect and evi can be further understood when considering that evi looks at anomalies and aspect mostly affects the overall biomass, not anomalies per se. Regarding the issue of scale sensitivity, we were only concerned with correlations at 0.1° resolution because that is the resolution at which the model was trained.

Line 207 – 212: Again a kind of repetioon of the goals. Could be considered included in the headings rather than repeating the goals for the third time.

This comment is in line with the reviewers' comments on 81-84. We kindly refer the reviewer to our responses to the reviewer's comments on lines 81-84.

Figure 3: What is stIl1.... same as stl1?

The authors thank the reviewer; this will be has been corrected. This was a typo and stll1 is meant to be read as stl1.

Line 221-222: This is not obvious from the figure 5 the reader do not know where urban and crop land is

We were meant to refer to Figure 6 only and have removed the erroneous reference to Figure 5.

Figure 4: Are there any system in where the model predicts better and worse? Places on earth where it is better or worse? Difficult to see this from the figure

We acknowledge that it is difficult to compare specific regions in Figure 4. We kindly refer the reviewer to Figures 5 and 6 which allow for spatial comparisons between different regions and landcover types. Additionally, since this comment is in line with the second comment, we refer the reviewer to our response there. In summary, we agree that additional emphasis is needed for comparisons between different regions. With the updated Figure 5, additional comparison between regions is possible.

Figure 5: A map showing were this are taken from is needed,

We acknowledge that this is an important addition and have included this in the revised manuscript.

Figure 6: Seems like mixed is poorer than crop. Urban is sign worse than the rest... why?

Given that this comment is in line with a following comment referring to line 289 (see below), we have given a detailed response to this issue there.

Line 233: What is ACC analysis

The ACC analysis refers to the anomaly correlation analysis which we introduce in Section 2.3 in the materials and methods. However, we have neglected to include the definition of the acronym. In the revised manuscript we have updated the sentence to read as follows: 'To determine to what extent this may influence the generality of the two products mentioned above, we further investigated the accuracy of the predicted evi under low growing conditions by calculating the anomaly correlation coefficient (ACC) index (Eq. 1)", line 215-217.

Line 247: Nowere discussed how good these input data are…. if the model is most sensitive to these, then this is important to discuss.

We kindly refer the reviewer to the first comment where we detail how we explain how we will add the important aspect of input data quality.

Line 256: 6% of what?… yes you mentioned before, but is it the same % of what. And is this a goal here to use as few data as possible in that case it would be valuable to see how more data influence

This will be clarified in the text and we mean 6% of all data available. To confirm, the goal is indeed to use as little data as possible because this in turn translates to the most efficient use of computational resources. In lines 192-197 we mention why we settled on a train:test ratio of 6%, however, we have aimed to reiterate this point by keeping the line 192-197 unchanged and providing additional context in the second paragraph of the materials and methods, lines 98-95.

Line 275: Are these errors more pronounced for some LC types or other paramenters lik hydrological and metorological input. Is the model underestimating or over estimatin?

We acknowledge that the manuscript refers to only the magnitude of errors but does not consider the direction. We have expanded on this detail in the results and pointed out for all landcover types the model fails to predicts extreme values. We have also highlighted that it under estimates evi in Urban areas, line 249.

Line 289: why? what features is not covered in your data for these areas. Harvesting is probably not covered by you input… is this an explanation…. …see this dicussion is coming later… but harvesting is not mentioned

We politely agree with the reviewer's insight and will include it in the revised manuscript. Furthermore, the reviewer previously stated that the mixed landcover type performs worse than the crop landcover type and poses the question why this may be. Given that the mixed landcover type includes Natural Vegetation and Crop mosaics, we infer that the reasons why mixed performs worse than croplands are very much in line with what the reviewer suggests, that we are missing features that are important for this landcover type. Additionally, since this landcover class consists of a mixture of two different landcover types that respond differently, it is plausible that the signal is relatively weak compared to other landcover types. We have added this rationale in the discussion, lines 330-337.

This study appears to focus on predicting global vegetation dynamics using the Enhanced Vegetation Index (EVI) and the Random Forest algorithm. However, the necessity and novelty of the study are unclear and poorly presented. There are several issues that need to be addressed in this revised version.

We want to thank the reviewer for reviewing our manuscript, the reviewer has provided us with valuable insights and suggestions that will improve the quality of this work. We have responded to the comments and suggestions in a point-wise fashion below.

1. The availability of 1km MODIS EVI through platforms like Google Earth Engine (GEE) is well-known. Therefore, the presented necessity of the study may be meaningless. This should be clarified in the introduction section.

We agree with the reviewer that there are various evi products available to users and acknowledge that the novelty and broader implications of this manuscript needs to be emphasized. Below, we detail how the crux of this manuscript extends beyond merely the description of a new data product.

Our focus was not to provide a new product to the already existing catalogue of evi (and other vegetation-related) datasets available on various platforms. Instead, the aim of this study was to investigate whether we could leverage a machine learning approach to overcome some of the limitations associated with the existing datasets the reviewer refers to. More specifically, already available datasets suffer from (a) limitations related to data gaps, (b) coarse temporal resolution originating from the creation of composite images, and (c) difference in spatial resolution when comparing older products to newer ones. In short, we aim to demonstrate that a fairly simple random forest algorithm can be used to (1) fill data gaps at resolutions that model is trained at and (2) downscale evi data to produce high-resolution data for the newer fine-scale evi data.

1) Gap-filling

Random forest based gap filling can be used to create temporally and spatially uniform data at a particular resolution in at least two ways: first, by filling data gaps originating from instances where data capture by satellites are interrupted by bad weather. Second, current evi products rely on the creation of composite images to produce spatially continuous data. For example, MODIS creates 16-day composites for terra and aqua satellites from various pass-over swathes. One of the implications from the current work suggests that RF can be trained on these incomplete swaths and then used to predict evi at the daily timestep; thereby producing continuous daily data.

2) Downscaling

Using the RF model to predict data at scales it was not trained on allows for the creation of long-term high-resolution evi products. For example, the already available MODIS 250m evi data set can be downscaled to spatial resolutions that are representative of the newer higher resolution products like Sentinel, which have a resolution of 30m. In this manuscript we show that scale commensurability between 10 km and 1km opens the door for the attempt described above.

Lastly, an important reason why we chose to predict values for an already existing dataset is that it allowed us to do a full validation on the results. Given all these points, we still acknowledge that this was not abundantly clear in the original version and have emphasised this in the abstract, introduction, and discussion in the revised manuscript.

2. There are numerous Random Forest-based approaches for gap-filling and downscaling satellite variables. It is not clear if the present study introduces any algorithmic improvements. These aspects are not adequately discussed in the introduction section.

The reviewer is correct to state that we did not provide any improvements to the RF regression algorithm. Instead we set out use this already widely used algorithm as a gap filling and downscaling tool that can be used to avoid some of the limitations associated with the current products. We have clarified this in the Introduction and Methods by highlighting the points we made in the previous comment.

3. The methodology section lacks detail. It appears that the present model establishes a Random Forest-based regression at a lower resolution scale and then applies it to a higher resolution scale. This process needs to be clearly explained in the manuscript.

We acknowledge that the manuscript in its current form does not articulate the aims and implications of our results to the broader HESS readership, and therefore we will aim to clarify these components in the revised manuscript. We have provided additional emphasis on this aspect of the methodology in the revised manuscript.. We have added additional context on line 84 -95 where we explain how input data at different resolutions were handled to produce the datasets used in this stud. We aimed to make it clear that, during the training step all data were provided at a 0.01° resolution. Whereas, during the prediction part of the process we matched all the inputs to the resolution we were predicting at. In other words, when predicting at the 0.01° resolution all input data were also at the 0.01° resolution and when predicting at the 0.001° resolution all input data were also at the 0.001° resolution.

4. The selection of feature variables is crucial in machine learning models. It is unclear why the author chose specific variables. Additionally, other variables such as surface temperature, precipitation, solar radiation, and surface albedo, which are known to be closely related to EVI, are not considered in this version. Furthermore, the low sharpness values of the drought index (Figure 3) may indicate a lack of relevance to EVI. However, the author insists on including this drought index as an essential feature.

We wanted to include variables that are representative of the water and energy balances experienced by vegetation. We briefly mentioned this rationale in lines 114-116. Regarding the variables that the reviewer suggests are important, we included precipitation (Table 1) and, as a proxy for surface temperature we also included soil temperature and two-metre temperature (t2m) in the initial manuscript. Therefore, we agree with the reviewer that these are important variables to consider. The rationale behind including the spei as input variable was to include a memory component into the model. We had not expected these to show strong influences but decided to include them given the importance of past conditions in determining current vegetation condition.

5. Most of the selected feature variables are derived from ERA products, which only provide a spatial resolution of approximately 10 km. Therefore, the applicability of using these variables at a 1 km scale is questionable.

This was one of the key tenants of this manuscript; in other words, can we provide RF with data at the coarse resolution so that it can accurately predict data at a finer resolution? Here, we show that this is possible, provided that one has some sort of ancillary fine-scale data to accompany the available coarse-scale data. We argue that even though one would expect that data at 10km to be too coarse to provide meaningful information our results suggest otherwise. In these results we show that, by relying on scale commensurability, we can use coarse-scale data to predict fine scale data. In addition, 10km resolution is the highest resolution currently available at the global scale, and studies like the one presented here are confined to use what is currently available. Using techniques, such as those presented in this study, that are capable of providing vegetation information at finer scales can open the door to more local assessments. We have added additional context in the throughout the introduction and have dedicated a section to how this applies to drought related studies in the discussion, lines 338-348.

6. Figures 6 and 7 do not provide sufficient information to demonstrate the merits of the predicted products in depicting drought. The context provided in sections 3.2 and 3.3 is insufficient.

We acknowledge that Figure 6 and 7 provide little information when making inference of the predictions in relation to drought. This is why, to our knowledge, we did not refer to Figures 6 or 7 when inferring the precision of the predictions in relation to drought. However, we do make reference to Figure 7 on line 233 to make the point that although these predictions are able to capture drought dynamics, it is to a lesser degree when compared to overall dynamics.

7. Finally, the author should compare the proposed model with existing models to highlight its novelty, particularly in the discussion section.

We agree that this will provide valuable insights and improve this manuscript; we have added a component on how our results relate with existing models and methods. To the best of our knowledge, there are currently no other models that employ this sort of approach for evi at the global scale. Given that this sort of comparison is vital, we have referred to a model that have employed this approach at the more regional level. For example, Roy (2021) provides an important comparison on the performance of different ML algorithms for predicting and forecasting MODIS ndvi and evi over Bangladesh. Comparing our results to the results of Roy (2021) will serve to place the current results within the wider framework of what has already been done. Given the lack of global studies related to our manuscript, we can only rely on studies that have employed a similar approach but with different variables. For example, Han et al. (2023) provide a good example how a similar approach to the one used here can produce global high resolution and continuous soil moisture. Please see lines 361 – 366 for the updated text.

References:

Bishal Roy. (2021) Optimum machine learning algorithm selection for forecasting vegetation indices: MODIS NDVI & EVI. Remote Sensing Applications: Society and Environment. Volume 23:100582.

Han Q, Zeng Y, Zhang L et al. 2023. Global long term daily 1 km surface soil moisture dataset with physics informed machine learning. Scientific Data 10:101.