Answers to Anonymous Referee #2

We thank the anonymous referee for reviewing the manuscript and for the valuable comments. We will revise the manuscript according to the suggestions. Below are the comments and our detailed responses.

Major Comment

Comment RC2.1 *The abstract lacks some context. Not many readers will know TSMP, so it needs to be pointed out more clearly (and earlier) what the applications of this model system are and, then, why it is important to generate pseudo observations of NDVI and BT from simulation output, when real Earth observations are available for these quantities. There are some hints at the end of the abstract (climate simulations), but this remains vague and doesn't help to understand why this study was performed. A side aspect of this is that this lack of claritiy makes it difficult to evaluate the stated model errors. Are MAEs of 0.027 for NDVI and 1.9 K for BT good or better than SOTA? What would be the reference here? Retrieval errors?*

Answer to RC2.1 We will revise the abstract accordingly. Please note that we do not aim to generate pseudo observations of NDVI/BT when satellite measurements are available but rather to predict NDVI/BT for periods where no satellite observations exist (future periods). However, we have to train and evaluate the model on a historical period. The real application is then to apply DL to a climate projection and predict NDVI/BT in the future. We agree that the stated model errors are hard to be interpreted without a reference. However, this is only mentioned to give an estimation about the accuracy that the model could achieve. The errors for satellite products highly depend on the spatial/temporal resolution and sensor being used. To the best of our knowledge, there are also no baselines models to compare with.

Comment RC2.2 *Section 1 takes a couple of short-cuts and doesn't always provide good explanations to motivate this study. Please see detailed comments below. Related to this, in section 2 it is unclear why, for example, radiative transfer models are discussed here, and some of the content of this section would better belong elsewhere.*

Answer to RC2.2 As suggested, we will revise section 1 and section 2.

Comment RC2.3 *Sections 3-6 are largely OK, except for minor comments listed below.*

Answer to RC2.3 We will revise sections 3-6 accordingly.

Comment RC2.4 *Section 7 variable importance: I like this analysis very much. However, I think one could point out that channel importance does not "explain" everything. First: if two variables are correlated, the network may decide to focus on one of them and the other one would seem unimportant, while it could provide almost the same information if it were alone. Second: seemingly "unimportant" variables may play an important role to get the final few percent accuracy out of the models. This could of course be tested by training a model on only the N most relevant variables and compare the results.*

Answer to RC2.4 We are aware that conclusions from such an analysis should be taken carefully as discussed in lines 654-663. We will move this discussion to Section 7.

Comment RC2.5 *Section 8: this discussion comes as a surprise as it goes much deeper into remote sensing and modelling issues than any of the other parts of the paper. As indicated above, there is lack of information in the Introduction and related work sections. I therefore suggest to re-arrange some of the text and use some of the material of the discussion in these earlier sections. The discussion could then be shortened and focus more on the applicability and prospects of the new method.*

Answer to RC2.5 As suggested, we will move parts of section 8 to earlier sections.

Minor Comments

Comment RC2.1 *Abstract l.3: why "intermediate step"? The image synthesis is the main product of the DL model, not an intermediate step in the modelling itself. The derivation of various indices is post-processing. Suggest to remove "in an intermediate step".*

Answer to RC2.1 We will remove it.

Comment RC2.2 *l.7: suggest rewording "... to assess the model's applicability to different seasons and regions..."*

Answer to RC2.2 We will revise it accordingly.

Comment RC2.3 *l.12 the unit of temperature is K, not K°.*

Answer to RC2.3 Thank you for this notice. We will check the manuscript and remove the symbol °.

Comment RC2.4 *Introduction l.20: Suggest to remove the first sentence (motherhood statement) and integrate "under a changing climate" in the following sentence, which provides a more concise and precise start of the text. Not all droughts are extreme, and while extreme events are a good motivation, this study does not focus on extreme events, but rather tries to provide information to assess droughts or the risk of droughts.*

Answer to RC2.4 We will revise the introduction accordingly.

Comment RC2.5 *l.32 delete "in the future"*

Answer to RC2.5 We agree as it looks redundant.

Comment RC2.6 *l.33 ff. The link made here between climate models and the water cycle is a bit too direct. It is a known weakness of general circulation models (aka "climate models") that convection and rainfall are not well captured over many world regions. This is why there is a need for more specialized hydrology models, which are frequently used for regional instead of global simulations. Also, the introduction of drought indices comes somewhat unmotivated. The rationale behind these is usually to convert information from some instrument (or model) into a meaningful* *quantity that can be used to assess the state of some ecosystem or the climate system. Why focus on agricultural indices here? If this is intended, then this should be stated in the first motivation sentences for this study.*

Answer to RC2.6 We will revise the introduction accordingly as this appears confusing. The study is primarily related to remote sensing based agricultural drought events.

Comment RC2.7 *l.50-55 The "discussion" about retrievals is good and can be used for the motivation of this study, but it is missing an explicit reference to retrieval errors. The problems with current retrievals (or perhaps even fundamental problems = theoretiscal limitations of physics-based retrievals?) should be described more precisely and with some more detail.*

Answer to RC2.7 We will revise the introduction to make it clearer. Please note that the study is not mainly related to retrieval errors.

Comment RC2.8 *l.66 I would avoid the word downstream-application here (even though it is technically correct) and rather formulate "To showcase the value (or potential) or our approach, we calculate (or derive) ..."*

Answer to RC2.8 We will revised it according to the suggestion.

Comment RC2.9 *l.74 This sentence is confusing in the context. Before, you give the impression that you rely on the model (implicitly assuming TSMP is perfect), whereas you now state that you can use the derived products to "examine the predicitve capability" of the model. As stated in the major comment above, the Introduction needs to be rewritten with a clearer explanation what this study is based on, how it is motivated (what doesn't work well at present?) and what are its primary objectives. Certainly, the aspect of model errors and their impact - or the potential of the method to quantify them - are one very relevant aspect that should at least shine through in the Introduction.*

Answer to RC2.9 Thank you for this comment. We will revise the introduction accordingly.

Comment RC2.10 *Section 2.1: the review of radiative transfer models is OK, but the reader doesn't understand why there is half a page or more on cloud retrievals when this paper is about vegetation indices. It would be helpful to add an introductional sentence or two explaining why section 2 is structured in the way it is and what content is expected. The discussion in 2.1 is perhaps a little too detailed.*

Answer to RC2.10 We will merge the first two sections and shorten the discussion.

Comment RC2.11 *l.118: The paragraph introducing the work of this study does not seem to connect to the general radiative transfer discussion above. The connection appears to be only methodologically (use of AI).*

Answer to RC2.11 Section 2 will be revised and integrated with the introduction.

Comment RC2.12 *l.132 grammar "the interaction ... exhibits ... behavior".*

Answer to RC2.12 We will correct it.

Comment RC2.13 *l.162 awkward phrasing "a single indicator like NDVI excluding BT" - do you mean "... either NDVI or BT"? Or simply cut after "indicator".*

Answer to RC2.13 We mean when only relying on NDVI. We will rewrite the sentence.

Comment RC2.14 *l.162 ff. After reading section 2, it becomes clearer what this paper aims to do. Some of the text here should be moved to the Introduction, and section 2 should no longer explain what is done in this study, but concentrate on discussing what has been made available so far.*

Answer to RC2.14 We will merge parts of section 2 with the introduction.

Comment RC2.15 *l.178 delete "at" before "IBG-3 institute".*

Answer to RC2.15 We will delete it.

Comment RC2.16 *l.180 "near nature realization" - what do you mean by this? Every model is an abstraction of some sort, and many models aim to produce realistic results. However, this expression is not scientific.*

Answer to RC2.16 We will remove it.

Comment RC2.17 *l.182 ff. please harmonize grammar in the bullet list - some bullets have verbs others don't.*

Answer to RC2.17 We will revise it.

Comment RC2.18 *l.190 "a dynamic equilibrium" - this is not unambiguous and depends on the choice of start and end date, for example.*

Answer to RC2.18 The dynamic equilibrium with the atmosphere (1979-1989) was obtained to initialize the subsurface and surface hydrologic and energy variables [1]. We will make this clearer in the revised manuscript.

Comment RC2.19 *l.195 comment, related to l.180: a free running model without DA will always be further away from "nature" than a model with DA.*

Answer to RC2.19 This is true. The simulation did not use any re-initialization or nudging. We will remove DA.

Comment RC2.20 *l.198 extending.*

Answer to RC2.20 It will be corrected.

Comment RC2.21 *l.199 why mention "with various vegetation types and climate conditions"? If you refer to Europe as a region, then this is kind of obvious and doesn't add information to* *this sentence describing the model set-up. If this refers instead to a property of the model or model output, then it doesn't belong here, but in a section where you describe the data and data distributions.*

Answer to RC2.21 We wanted to emphasize that the extension of the region we study (Europe) includes different vegetation types and climate conditions. We agree this seems obvious and we will delete it.

Comment RC2.22 *l.201 and *the* model set-up.*

Answer to RC2.22 It will be corrected.

Comment RC2.23 *l.204 I think you could add an extra sentence to say that DL has already been applied to TSMP simulations, instead of just referring to the papers.*

Answer to RC2.23 We will revise it accordingly.

Comment RC2.24 *l.221 what are upper and lower bounds of an ecosystem? Do you mean bounds of NDVI and BT for a specific ecosystem class? Also, the sentence "Consequently, ..." doesn't fit well. Better to write "Hence, ..." or "Thus, ..." or "Therefore, ..."*

Answer to RC2.24 The upper and lower bounds are the min and max values for NDVI and BT for a specific pixel. We will make this clearer and rephrase it.

Comment RC2.25 *l.230 Is alpha a fixed coefficient or does it vary with ecosystem class or other parameters?*

Answer to RC2.25 In our work, we used a standard value $\alpha = 0.5$ [2]. Please note that this does not affect NDVI and BT predictions. As mentioned in lines 267-269 and 637-639, it is better to calibrate α with respect to the location (see [3, 4]). However, this is beyond the scope of our work to improve the weighting.

Comment RC2.26 *l.234 delete "Moreover,".*

Answer to RC2.26 It will be deleted.

Comment RC2.27 *l.248-254 grammar (OK, but clearly non-native English).*

Answer to RC2.27 We will proofread the section.

Comment RC2.28 *l.271 Please describe the data cube dimensions. Is there one datacube with (time, lat, lon) for each variable? Remove "observed".*

Answer to RC2.28 We stored a datacube for each week (variable, lat, lon). We will clarify this. "observed" will be removed.

Comment RC2.29 *l.275 zero*

Answer to RC2.29 It will be corrected.

Comment RC2.30 *l.282 the theta doesn't belong in eq 7, which describes the mapping objective. It only comes in when you in fact use a model, i.e. when you describe the U-net.*

Answer to RC2.30 We will reorder the text and put eq 7 after we mention the model.

Comment RC2.31 *l.300 number of channel*s*.*

Answer to RC2.31 It will be corrected.

Comment RC2.32 *l.325 period missing. And: pixel representations.*

Answer to RC2.32 It will be corrected.

Comment RC2.33 *l.328 Please be more specific: you refer to the quadratic scaling, which primarily limits the attention span, but not "applications" per se.*

Answer to RC2.33 You are right. We refer to the quadratic computation complexity of the self-attention. We will make this clearer.

Comment RC2.34 *l.337 input channel*s*.*

Answer to RC2.34 It will be corrected.

Comment RC2.35 *l.347 remove "a" before "one".*

Answer to RC2.35 It will be corrected.

Comment RC2.36 *l.350 reduce the number of model parameters.*

Answer to RC2.36 We will revise it as suggested.

Comment RC2.37 *eq.11 I suggest to replace the somewhat clumsy expressions FocalModulationBlock etc. by shorter variable names which then need to be defined in the text, of course. This would improve readability of the equation.*

Answer to RC2.37 We will simplify the terms.

Comment RC2.38 *l.379 "less" compared to what? I assume you mean MSE loss.*

Answer to RC2.38 Correct. We mean here less than MSE. We will make this explicit in the revised manuscript.

Comment RC2.39 *l.392 play *a* more important role.*

Answer to RC2.39 It will be corrected.

Comment RC2.40 *l.425 I think it would be easier to describe the U-net baseline model by simply stating what it consists of instead of "reverse engineering" it by abstracting the focal attention* *blocks away.*

Answer to RC2.40 We will revise the description.

Comment RC2.41 *l.428 ff Please provide a few more details on the competitor models, such as number of layers, size of attention matrix etc. This could also be summarized in an Appendix, which should then be referenced here.*

Answer to RC2.41 All baseline models share the number of layers but differ in the type of layers. The U-Net model does not use an attention mechanism, but follows the original U-Net design using 2D convolutions. Swin Transformer uses self-attention inside local windows. The implementation details about the models are mentioned in section 5 "Implementation details".

Comment RC2.42 *l.438 Do you mean "Apart from"?*

Answer to RC2.42 Yes. We will correct it.

Comment RC2.43 *l.442 Why is the second climatology computing the future? 2016 is in the past. Also, grammar: future should be singular.*

Answer to RC2.43 The main application of the study is to estimate the vegetation condition using a DL-based model for future periods where no satellite images are available. We compare the results of the DL models to two NDVI/BT climatologies from remote sensing observations. The first climatology (1981-1988) is used to show the limitation of using prescribed satellite phenology for future projection and we argue that a DL-based model maybe a better replacement (for example future simulations often prescribe vegetation condition in a satellite phenology-mode neglecting the inter-annual variability [5]). This is still not a fair comparison because the DL-models were trained on periods succeeding this climatology. The second climatology (1989-2016) is used to show that the DL-models still outperform the climatology. We could compute the second climatology (1989-2016) because we used a climate simulation in an overlapping period with historical remote sensing observations. This comparison is not possible for future climate projections since no satellite observations exist yet. We will rewrite lines 438-444 to make it clear.

Comment RC2.44 *l.450 randomly perturbing.*

Answer to RC2.44 It will be corrected.

Comment RC2.45 *l.451 this seems to contradict the preprocessing description in section 3. There you wrote that samples were averaged over a week, which implies that *all* samples are used in thre average. What is written here is a random estimator of the weekly average based on two days.*

Answer to RC2.45 We use this as an augmentation technique only during training. For validation and test, we average all days. We will make this clearer in the revised manuscript.

Comment RC2.46 *l.460 typo finally.*

Answer to RC2.46 It will be corrected.

Comment RC2.47 *l.466 remove comma.*

Answer to RC2.47 It will be corrected.

Comment RC2.48 *l.475 I don't understand the reference to radiative transfer models here and suggest it be removed.*

Answer to RC2.48 We will remove it as suggested.

Comment RC2.49 *l.484 As shown.*

Answer to RC2.49 It will be corrected.

Comment RC2.50 *l.486 shown *in* Figs...*

Answer to RC2.50 It will be corrected.

Comment RC2.51 *l.493 weakness*es*.*

Answer to RC2.51 It will be corrected.

Comment RC2.52 *l.505 This discussion on errors could go one step further. While I agree that a full error attribution may be beyond the scope of this paper, you could at least give some indications, which errors come from the TSMP input data, which are from the observations and which may be DL model errors. This would be especially relevant for the TSMP data. For example, by showing the variability in different 2-day "weekly" samples and how this translates into different DL model outputs. Or you could apply some systematic perturbations to the TSMP output, thereby roughly correcting known model biases, and then see how this change the results. (OK, l.537 provides at least some indication already, and you could also refer to section 7 for additional insights).*

Answer to RC2.52 We will extend the discussion and refer to section 7 as suggested.

Comment RC2.53 *l.516 typo "for".*

Answer to RC2.53 It will be corrected.

Comment RC2.54 *l.540 , who showed.*

Answer to RC2.54 It will be corrected.

Comment RC2.55 *l.541 affecting.*

Answer to RC2.55 It will be corrected.

Comment RC2.56 *l.723 Klaus Goergen.*

Answer to RC2.56 It will be corrected.

References

[1] C. Furusho-Percot, K. Goergen, C. Hartick, K. Kulkarni, J. Keune, and S. Kollet, “Pan-european groundwater to atmosphere terrestrial systems climatology from a physically consistent simulation,” *Scientific Data*, vol. 6, no. 1, p. 320, 2019. [Online]. Available: https://doi.org/10.1038/s41597-019-0328-7

[2] W. Yang, F. Kogan, and W. Guo, “An ongoing blended long-term vegetation health product for monitoring global food security,” *Agronomy*, vol. 10, no. 12, 2020. [Online]. Available: https://www.mdpi.com/2073-4395/10/12/1936

[3] J. Zeng, R. Zhang, Y. Qu, V. A. Bento, T. Zhou, Y. Lin, X. Wu, J. Qi, W. Shui, and Q. Wang, “Improving the drought monitoring capability of vhi at the global scale via ensemble indices for various vegetation types from 2001 to 2018,” *Weather and Climate Extremes*, vol. 35, p. 100412, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212094722000068

[4] J. Zeng, T. Zhou, Y. Qu, V. Bento, J. Qi, Y. Xu, Y. Li, and Q. Wang, “An improved global vegetation health index dataset in detecting vegetation drought,” *Scientific Data*, vol. 10, p. 338, 05 2023.

[5] D. M. Lawrence, R. A. Fisher, C. D. Koven, K. W. Oleson, S. C. Swenson, G. Bonan, N. Collier, B. Ghimire, L. van Kampenhout, D. Kennedy, E. Kluzek, P. J. Lawrence, F. Li, H. Li, D. Lombardozzi, W. J. Riley, W. J. Sacks, M. Shi, M. Vertenstein, W. R. Wieder, C. Xu, A. A. Ali, A. M. Badger, G. Bisht, M. van den Broeke, M. A. Brunke, S. P. Burns, J. Buzan, M. Clark, A. Craig, K. Dahlin, B. Drewniak, J. B. Fisher, M. Flanner, A. M. Fox, P. Gentine, F. Hoffman, G. Keppel-Aleks, R. Knox, S. Kumar, J. Lenaerts, L. R. Leung, W. H. Lipscomb, Y. Lu, A. Pandey, J. D. Pelletier, J. Perket, J. T. Randerson, D. M. Ricciuto, B. M. Sanderson, A. Slater, Z. M. Subin, J. Tang, R. Q. Thomas, M. Val Martin, and X. Zeng, “The community land model version 5: Description of new features, benchmarking, and impact of forcing uncertainty,” *Journal of Advances in Modeling Earth Systems*, vol. 11, no. 12, pp. 4245–4287, 2019. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001583