Answers to Anonymous Referee #1

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 RC1.1 *The introduction lacks rationales on conducting the study. I suggest integrating sections 2.1 and 2.2 into the introduction. In the introduction, I would like to emphasize that the authors need to summarize what previous studies did, not describe what they did one by one (first paragraph in section 2.1 and first paragraph in section 2.2, which are not very readable and not necessary), but summarize and provide rationales and research gaps that this study will fill.*

Answer to RC1.1 We will revise the introduction and integrate it with the related works section.

Comment RC1.2 *Lines 190-192: what is the rationale on variable selection?*

Answer to RC1.2 Only variables within the period applicable for analysis were selected. Some variables were not selected due to the reason that for some variables in the model, the full groundwater states equilibrium starts after 1995. In addition, we did not select the variables that do not cover the whole period of the simulation.

Comment RC1.3 *Section 4.1: the authors need to provide rationale and advantages about the proposed architecture at the beginning of this section compared to state of art architectures (information in lines 322 to 333 with adjustments should be provided at the beginning of section 4.1). I also suggest using jargons as less as possible in the methodology section.*

Answer to RC1.3 We will revise section 4.1 accordingly and simplify some terms when possible.

Comment RC1.4 *I suggest that lines 403-464 should be in the methodology section as baseline approaches and evaluation metrics. I also suggest the italics section in section 5 to be separated section as 5.1, 5.2 and so on, which will be much clear.*

Answer to RC1.4 We will revise the manuscript accordingly.

Comment RC1.5 *It is not clear what the authors would like to emphasize in lines 438-444.*

Answer to RC1.5 The main application of the study is to estimate the vegetation condition using a DL-based model for periods where no satellite images are available. As a baselines, we compare the results of DL models to two NDVI/BT climatologies from remote sensing observations (the climatology is computed pixel-wise and on a weekly basis). The climatology has the limitation that the inter-annual variability in NDVI/BT is neglected as average cycles are used. The first climatology (1981-1988) is used to show the limitation of using prescribed satellite phenology for future projection and we argue 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 [1]). This is still not a fair comparison because the DL-models were trained on periods succeed this climatology. The second climatology (1989-2016) is used to show that the DL-models still outperform the climatology and generalized beyond the mean annual cycles of NDVI/BT. We could compute the second climatology (1989-2016) because we used a climate simulation in an overlapped period with historical remote sensing observations. This comparison is not possible for future climate projections since no satellite observations exist yet. Please note that we excluded the validation/testing (2010-2012) years for the computation of this second climatology. The last part is that we could not compare to a specific physically-based model since such a model has to construct the albedo and emission for visible, near-infrared, and infrared bands under all conditions and for all surface types and then trace the atmospheric correction (NOAA AVHRR depends on temporal compositing for the atmospheric correction) to the top of the atmosphere regardless of the solar and satellite geometry. We will rewrite lines 438-444.

Comment RC1.6 *It is not clear in line 451. Based on the methodology (lines 285-289), both the input variables and outputs are in weekly time scale, why here claims average two days?*

Answer to RC1.6 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 RC1.7 *The results are not well organized and too many sections. I suggest organizing all the results into one section as subsections for each topic*

Answer to RC1.7 We will revise the manuscript accordingly.

Comment RC1.8 *Lines 467-471: it is confusing about different climatology. Based on the dataset information in section 3.1, the model simulation is from 1989 to 2019. What does climatology from 1981 to 1988 come from? What is the main point for analyzing different climatology (climate change?)? Table 1 is also confusing for mixing climatology different periods and validation/test different years.*

Answer to RC1.8 Please see the answer to RC1.5. We agree that the comparison to climatology in tables may appear confusing. We will make it clearer in the suggested new section "baseline approaches" and change the name of climatology (1981-1988) to climatology I and climatology (1989-2016) to climatology II. For validation/test, we took the last 3 years of AVHRR (2010-2012) and the last 3 years of converted VIIRS (2017-2019).

Comment RC1.9 *Based on Table 1, the metric differences for different DL models are trivial (mostly around percentage scale), although the DL models have very different building blocks. The proposed DL model is also not consistent better than others particularly for the test dataset. My doubt is whether the trivial differences are caused by stochasticity not due to DL model itself. The authors claimed fixed random seed for reproductivity, but what if the seed is not fixed and what does the results look like for running the model several times? Will the results be consistent with any conclusion got from the table?*

Answer to RC1.9 Table A reports the mean and standard deviation for training with 3 different random seeds and we will replace Table 1. In general, there are some differences for the results on the test set and we will adjust our comments accordingly in the revised manuscript. We would also like to note that all of the described models are our implementation and we found Focal Modulation among them to work overall the best. We also replaced the name 2D CNN with U-Net to be consistent through out the manuscript.

Table A: Comparing the performance of different DL models on the validation and test sets.

Algorithm	MAE(↓)	RMSE(↓)	R^2(↑)	R_p(↑)	R_s(↑)
Validation - Years (2010, 2011, 2017) - 156 weeks					
NDVI					
Climatology 1981-1988	0.0550	0.0680	0.5763	0.8939	0.8669
Climatology 1989-2016	0.0326	0.0416	0.8372	0.9353	0.9113
U-Net	0.0277 ±0.0001	0.0365 ±0.0002	0.8743 ±0.0008	0.9406 ±0.0005	0.9172 ±0.0005
Wave-MLP	<u>0.0272</u> ±0.0003	**0.0358** ±0.0003	<u>0.8784</u> ±0.0018	<u>0.9422</u> ±0.0018	<u>0.9183</u> ±0.0021
Swin Transformer V1	0.0273 ±0.0003	0.0362 ±0.0003	0.8759 ±0.0022	0.9411 ±0.0013	0.9161 ±0.0023
Swin Transformer V2	0.0277 ±0.0003	0.0369 ±0.0003	0.8703 ±0.0021	0.9415 ±0.0010	0.9167 ±0.0008
Focal Modulation	**0.0269** ±0.0001	**0.0358** ±0.0002	**0.8790** ±0.0017	**0.9432** ±0.0001	**0.9194** ±0.0009
BT (K)					
Climatology 1981-1988	2.9130	3.7302	0.8454	0.9466	0.9408
Climatology 1989-2016	2.3017	3.0020	0.8963	0.9601	0.9539
U-Net	1.9377 ±0.0093	2.6067 ±0.0057	0.9243 ±0.0014	0.9667 ±0.0004	<u>0.9603</u> ±0.0007
Wave-MLP	<u>1.9200</u> ±0.0491	<u>2.5834</u> ±0.0486	<u>0.9248</u> ±0.0035	<u>0.9668</u> ±0.0006	<u>0.9603</u> ±0.0007
Swin Transformer V1	1.9642 ±0.0246	2.6341 ±0.0303	0.9221 ±0.0012	0.9661 ±0.0005	0.9590 ±0.0006
Swin Transformer V2	1.9741 ±0.0191	2.6420 ±0.0258	0.9225 ±0.0013	0.9659 ±0.0011	0.9590 ±0.0014
Focal Modulation	**1.9010** ±0.0071	**2.5364** ±0.0073	**0.9280** ±0.0012	**0.9679** ±0.0001	**0.9614** ±0.0007
Test - Years (2012, 2018, 2019) - 139 weeks					
NDVI					
Climatology 1981-1988	0.0567	0.0697	0.5529	0.8933	0.8704
Climatology 1989-2016	0.0314	0.0400	0.8507	0.9433	0.9254
U-Net	0.0274 ±0.0004	0.0359 ±0.0005	0.8772 ±0.0006	0.9435 ±0.0006	0.9237 ±0.0009
Wave-MLP	**0.0261** ±0.0006	**0.0343** ±0.0008	**0.8861** ±0.0043	**0.9467** ±0.0024	<u>0.9252</u> ±0.0011
Swin Transformer V1	0.0269 ±0.0003	0.0355 ±0.0004	0.8795 ±0.0029	0.9442 ±0.0010	0.9239 ±0.0014
Swin Transformer V2	0.0270 ±0.0005	0.0359 ±0.0005	0.8766 ±0.0038	0.9447 ±0.0012	0.9251 ±0.0020
Focal Modulation	<u>0.0266</u> ±0.0003	<u>0.0350</u> ±0.0004	<u>0.8808</u> ±0.0014	<u>0.9454</u> ±0.0009	**0.9253** ±0.0016
BT (K)					
Climatology 1981-1988	2.8806	3.6864	0.8447	0.9485	0.9470
Climatology 1989-2016	2.2024	2.8880	0.9036	0.9623	0.9606
U-Net	1.9920 ±0.0148	2.6652 ±0.0262	0.9164 ±0.0021	0.9644 ±0.0009	0.9616 ±0.0005
Wave-MLP	<u>1.9376</u> ±0.0184	<u>2.6221</u> ±0.0177	0.9172 ±0.0005	0.9647 ±0.0005	0.9619 ±0.0008
Swin Transformer V1	1.9563 ±0.0329	2.6381 ±0.0397	0.9169 ±0.0038	<u>0.9649</u> ±0.0009	<u>0.9627</u> ±0.0008
Swin Transformer V2	1.9516 ±0.0639	2.6277 ±0.0874	<u>0.9183</u> ±0.0060	0.9641 ±0.0025	0.9619 ±0.0020
Focal Modulation	**1.9179** ±0.0458	**2.5745** ±0.0470	**0.9204** ±0.0030	**0.9664** ±0.0007	**0.9636** ±0.0006

Comment RC1.10 *I suggest combining section 8 and 9 as one section (conclusions and limitations). The current discussion section appears only limitations not discussion.*

Answer to RC1.10 We will combine the sections.

Minor Comments

Comment RC1.1 *At the beginning of abstract, I suggest adding one or two sentences about research problem the authors would like to solve.*

Answer to RC1.1 We will revise the abstract accordingly.

Comment RC1.2 *AVHRR in line 44 firstly appear without full name and also check for other short names.*

Answer to RC1.2 We will check the abbreviations and adjust the manuscript accordingly.

Comment RC1.3 *Line 50: change synthesis to be 'synthesize'.*

Answer to RC1.3 We will correct it.

Comment RC1.4 *Line 195: What is DA represented? If the term only shows once in the manuscript, it is not necessary to use the short name.*

Answer to RC1.4 DA refers to data assimilation. We will replace DA by the full name.

Comment RC1.5 *Line 200: theta and lambda represents?*

Answer to RC1.5 Theta and Lambda refer to the coordinates of the rotated pole. We will rewrite the terms as follows: "... The grid specification for TSMP is a standardized rotated coordinate system ($\phi_{(rotated\ pole)} = 39.5°$ N, $\lambda_{(rotated\ pole)} = 18°$ E) with a spatial resolution ..."

Comment RC1.6 *Line 253: Is it equation (5) is for NDVI for AVHRR instead of VIIRS as stated in line 253?*

Answer to RC1.6 Thank you for this notice. Equation (5) computes NDVI for the AVHRR system. Lines 252-253 should be "... re-composite NDVI for AVHRR ...". We will correct it in the revised manuscript.

Comment RC1.7 *Figure 1: It is better to use TSMP instead of TerrSysMP since TSMP is used all over the manuscript.*

Answer to RC1.7 We will change the figure accordingly.

Comment RC1.8 *Line 325: it should have a period between 'efficiently In'.*

Answer to RC1.8 We will correct it.

Comment RC1.9 *Line 368-369: it is not clear why the third gate focuses on the water area? How do you know that?*

Answer to RC1.9 The bright colors for specific regions in Fig. 3 represent higher values, which correspond to higher attentions of the model to that regions. For G_3^k, the attention values are high at areas with water. We will revise lines 368-369 to make it clear.

Comment RC1.10 *Add references on line 430.*

Answer to RC1.10 We will add the following references [2–5].

Comment RC1.11 *Line 448-450: the sign N(0, I =0.02) is not clear. It is better to use normal distribution with zero mean and standard deviation.*

Answer to RC1.11 We will replace the terms as suggested.

Comment RC1.12 *Line 461: reported how long it takes for the focal model, what about other DL models?*

Answer to RC1.12 In Table B, we report the inference time to generate one sample for NDVI and BT containing 397×409×2 grid points. We will include the table.

Table B: Inference time in seconds for different DL models.

Algorithm	GPU[1]	CPU[2]
U-Net	0.09 ±0.02	5 ±0.2
Wave-MLP	0.28 ±0.00	10 ±0.3
Swin Transformer V1	0.18 ±0.00	11 ±0.2
Swin Transformer V2	0.19 ±0.00	11 ±0.2
Focal Modulation	0.24 ±0.01	12 ±0.1

[1]NVIDIA GeForce RTX 3090 GPU
[2]AMD Ryzen 9 3900X 12-Core CPU

Comment RC1.13 *Line 471-472: What is these non-ML baselines? It is not clear for this statement.*

Answer to RC1.13 The non-ML baselines represent the climatology. We will rewrite the sentience to make it clearer in the revised manuscript.

Comment RC1.14 *Line 516: change 'fro' to 'for'.*

Answer to RC1.14 We will correct it.

Comment RC1.15 *Figure 10: change 'sprint' to 'spring' in the left two plots in the first row.*

Answer to RC1.15 We will correct it.

References

[1] 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

[2] L. Wang, R. Li, C. Duan, C. Zhang, X. Meng, and S. Fang, "A novel transformer based semantic segmentation scheme for fine-resolution remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[3] L. Gao, H. Liu, M. Yang, L. Chen, Y. Wan, Z. Xiao, and Y. Qian, "Stransfuse: Fusing swin transformer and convolutional neural network for remote sensing image semantic segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 10 990–11 003, 2021.

[4] L. Wang, R. Li, C. Zhang, S. Fang, C. Duan, X. Meng, and P. M. Atkinson, "Unetformer: A unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 190, pp. 196–214, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271622001654

[5] A. A. Aleissaee, A. Kumar, R. M. Anwer, S. Khan, H. Cholakkal, G.-S. Xia, and F. S. Khan, "Transformers in remote sensing: A survey," *Remote Sensing*, vol. 15, no. 7, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/7/1860