Dear authors,

After the previous round with substantial revisions, reviewer #1 is satisfied however due to the strong disagreements between the two original two reviewers I asked another expert to have an additional look and he find that major revision are required. Please address these and I'll share the manuscript only with this new reviewer for a review.

Kind regards,
Niko

> Response: Thank you, Dr. Wanders and the first two reviewers for helping to improve our manuscript. We detail responses to the additional reviewer below.

Reviewer 1: This is a revision of a previous submission which required additional work before publication. The authors have done a good job addressed the comments following the previous review. In particular, the model comparisons, figures and discussion/conclusions have been improved. Hence, I would recommend publication in HESS.

Reviewer 2:
Dear authors,

Please find attached an annotated PDF with my comments. Briefly summarizing those:

1. I would strongly recommend you clarify the novelty and/or relevance of the analysis you present. The main conclusion seems theoretically clear to me: land cover classifications can be inaccurate, and using inaccurate land cover classifications as model inputs leads to unhelpful model results. I don't think anyone will argue with this, but it is somewhat unclear to me what your analysis adds over what is already known about this problem.

> Response: Thank you, Dr. Knoben, for the thorough review and improvements to the clarity and impact of our manuscript. We have updated the manuscript to include more details about the novelty of our findings and technical workings of the models. We detail these in responses to your major comments below.
>
> In summary, hydrologic and water quality models, such as the one used in the manuscript, are traditionally developed using LULC classified during the growing season or annually. Previous work has largely focused on impacts of LULC changes on these models over longer (e.g., 10+ year) time spans. However, the recent availability of analysis-ready LULC at higher spatiotemporal resolution such as Dynamic World has made it more feasible for models to incorporate non-growing season LULC or LULC updates at shorter time spans (for instance, simulating the impacts of new settlements or

forest conversion on river dynamics and biogeochemical budgets). This makes evaluations of seasonal LULC classification impacts on models critical as we move into higher spatiotemporal resolution for study designs and interpretations of results.

So far as we know, ours is the first study to evaluate the use of the seasonal Dynamic World data in a hydrologic and water quality modeling framework. Our study is novel in that we provide guidance for other modelers who are aiming to use the high spatiotemporal resolution LULC data, using SWAT or other models. We have now updated our introduction and discussion to add clarity for the novelties, and provided additional beneficial guidance following your ideas. Page and line numbers in our responses represent locations in the track changes manuscript.

2. I believe the description of how HRUs are defined needs more work. This is important because the difference in LULC are propagated onto the different HRUs. Knowing more clearly which data fed into the HRUs, and how the size and location of the HRUs compares to the size and location of the LULC changes/uncertainties you identify is critical for correctly interpreting the results shown in this paper.

Response: To aid with interpreting our results, we have now included additional descriptions for how precisely HRU LULC characteristics were distributed in each modeling case, including maps and charts to help visualize the sizes and locations of HRUs in the models (see below for Figures S4 and S5). We also expand on why HRU delineation is pertinent for our study, as it is a way by which seasonal variation in LULC estimates affects model responses. We describe results for where specific HRU characteristics are being impacted by the model's LULC input, and use the additional HRU detail to further develop guidance for using high spatiotemporal frequency LULC data in hydrologic and water quality models. In particular, we discuss the different HRU delineation strategies such as aggregating by dominant LULC characteristics or maximizing HRU resolution, and how each could be affected by LULC seasonality, while also introducing your idea for separating stable from changing HRUs. Full descriptions for these actions are in the responses below.

Kind regards,
Wouter Knoben

Responses to PDF comments:

1 Introduction. SUMMARY:

I think the aspects that are covered int his introduction make sense, but what I think is lacking is an overview of the current understanding of this topic. What is already known about the use of long-term average LULC data vs seasonal LULC data in modelling contexts?

It may make sense to also add on the topic of uncertainty in this literature overview, particularly because some products (e.g., MODIS: see quote below) specifically recommend that users do not use the year-to-year variation in LULC maps due to issues with the classification algorithm. What does this imply for the use of LULC maps at shorter temporal time scales?

--
"After stabilization, the classifications are condensed into the final set of six legends and associated QA information. Despite improving the stability to the product, we urge users not to use the product to
determine post-classification land cover change. The amount of uncertainty in the land cover labels for any one year remains too high to distinguish real change from changes between classes that are spectrally indistinguishable at the coarse 500-m MODIS resolution"

https://lpdaac.usgs.gov/documents/101/MCD12_User_Guide_V6.pdf

> Response: We have added further literature review to our introduction that includes an overview of temporal LULC inconsistencies and approaches to address them, as well as transitioning to how hydrologic and water quality models fit in.
>
> "*Addressing temporal inconsistencies is important for accurately identifying LULC change (Sexton et al., 2013; Liu and Cai, 2012; Hermosilla et al., 2018) and various approaches have been developed that include incorporating time as a co-dependent in the classifier to remove illogical changes (Graesser et al., 2022), and probability-based statistics to separate noise from trends (Zhu et al., 2012; Zhu and Woodcock, 2014; Sulla-Menashe et al., 2019; Zhao et al., 2019). However, these approaches are typically not readily incorporated into watershed-scale hydrologic and water quality model frameworks, which take pre-classified LULC as model input (Li et al., 2019).*" (page 2, lines 41-46 in the track changes manuscript)

Page 2, lines 59-60. Note to self: impact of parameter differences on long-term projections could be relevant.

2 Materials and Methods. SUMMARY:
This section is mostly complete but various clarifications are needed. In particular the way in which HRUs are defined is critical to understanding the study (because the impact of seasonal LULC changes affects the dominant land cover on a per-HRU basis), and needs to be clarified.

Response: We have expanded our descriptions of model HRU assignments and also provided figures to visualize the approach. Further descriptions of how HRU LULC populations affect modeling results can be found in our response to comment "Page 6, lines 127-128." below.

"*SWAT divides a watershed into spatial subbasins, which may be further divided into unique combinations of soils, landuse, and slopes called Hydrologic Response Units (HRUs). HRUs are pertinent to this work as their delineations are in part determined by LULC. HRUs are a thus a mechanism by which differences in LULC classification, including erroneous differences derived from seasonality in remote sensing data, can impact the model. Subbasins were delineated using the program QSWAT. In the development of the SWAT models, one spatial data layer for each of elevation, soils, and LULC was input to generate tables that represent base watershed conditions.*" (page 6, lines 141-147)

"*The Rock Creek models for LULC change simulation (Case #2) had 13 subbasins, each assigned the dominant HRU, as has been done to more efficiently use computational resources (Myers et al., 2021b; Arabi et al., 2008). Gridded 4 km GridMET historic weather inputs were used as the Rock Creek watershed extends over 30 km from north to south (Abatzoglou, 2013). The Difficult Run SWAT models (Case #3) had 7 subbasins. Our Difficult Run Watershed SWAT models were constructed so that the maximum number of HRUs was incorporated (i.e., no minimum HRU area threshold), as has been done to compare independently calibrated model performance (Fuka et al., 2012), with weather data from National Oceanic and Atmospheric Administration (NOAA) station USW00093738 (Table S1). Further descriptions of model HRU numbers and proportions of watershed HRU areas with different LULC inputs can be found in Figures S4 and S5.*" (page 6, lines 149-156)

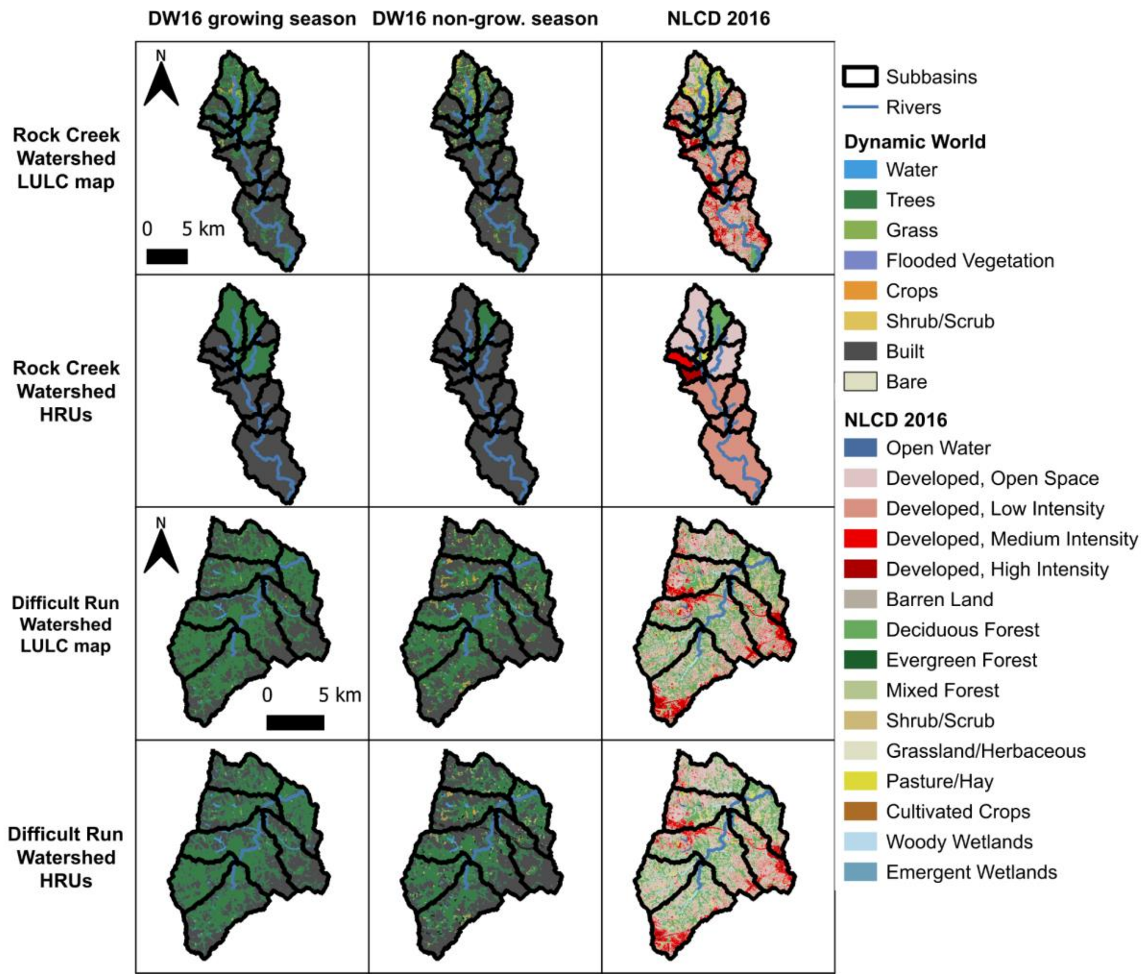

Figure S4. Delineated hydrologic response units (HRUs) for the Rock Creek Watershed (Case #2; top two rows) and Difficult Run Watershed (Case #3; bottom two rows). Rows show the input LULC data for each watershed and the resulting HRUs, while columns differentiate Dynamic World 2016 (DW16) growing and non-growing seasons and NLCD 2016. HRU numbers for Rock Creek were 13 for each LULC input, as each subbasin was assigned the dominant combination of LULC, soils, and slopes. HRU numbers for Difficult Run (which used the maximum HRU number approach) were 43 for Dynamic World 2016 growing season, 48 for Dynamic World 2016 non-growing season, and 111 for NLCD 2016 input.

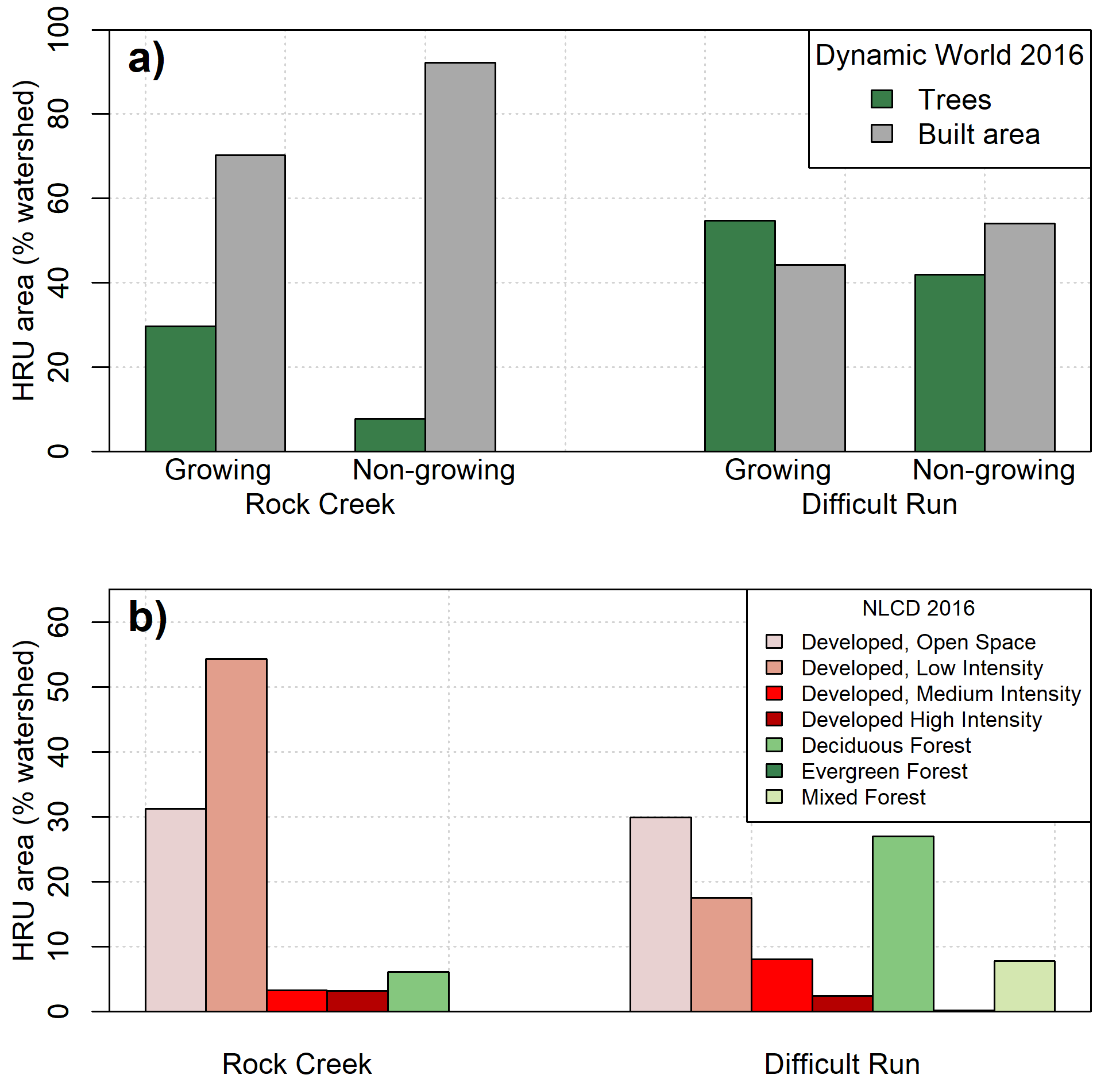

Figure S5. a) Proportions of HRU area being populated with Dynamic World 2016 trees and built area classes for the Rock Creek and Difficult Run Watersheds, split between growing and non-growing season inputs, and b) Proportions of HRU area being populated with NLCD 2016 developed and forest classes. All other HRU LULC assignments combined made up 0-4% of each watershed for Dynamic World inputs, and 2-7% of each watershed for NLCD inputs. For Rock Creek (dominant HRU approach), there was a difference in HRUs populated with Dynamic World 2016 trees class of 21.8% between growing and non-growing seasons, while that difference was 12.8% for Difficult Run (maximum HRU approach).

Page 3, lines 79-80. It's not (yet) clear to me why this sentence is relevant to the current manuscript.

<u>Response</u>: We have reworded this sentence to be clearer.

> "*Median SC values over the entire time period were used as the indicator of anthropogenic impacts to each stream for comparisons between monitoring sites (Dow and Zampella, 2000).*" (page 4, lines 91-93)

Page 4, lines 85-87. I'm uncertain yet if this will matter, but this feels a little odd to me. I can envision a case where, due to seasonal changes in vegetation density, the dominant land class changes from vegetated to barren (or some sort of similar change). It's only a limitation of SWAT that this would represent such a sharp change in model configuration. Other models can have the ability to specify different HRUs for different land classes (e.g. SUMMA), or use internal tiling to account for LULC variations (e.g. VIC, MESH).

Can the authors clarify if this investigation is specifically targeted at SWAT or if these findings are applicable to a wider array of models/scenarios?

Note: upon reading further Fig. 3a shows exactly this situation. Some discussion is warranted.

> <u>Response</u>: In our discussion, we describe the applicability of this approach to different LULC differentiation approaches including tiling and HRUs (used by SWAT), albeit we had not yet introduced these concepts at this point. In short, the impacts of illogical LULC changes are pertinent to these different models and not limited to SWAT, since they become simulated as an actual LULC change, which would affect model outputs. We updated our text here so not to appear overly-targeted to SWAT early on. We also further discuss modeling HRU approaches in our response to the comment below beginning with "Page 9, lines 204-209."
>
> "*Thus, there was one composite image for each season (growing and non-growing) that represented the most common LULC class for each pixel over the time period of individual images, to input into the hydrologic and water quality models*." (page 4, lines 98-100)
>
> "*Illogical LULC classifications related to seasonal differences in remote sensing data could be pertinent to models beyond our cases of regressions and SWAT in the eastern United States, such as models for which accurate parameterization of LULC processes is essential for simulating the impacts of climate change (Glotfelty et al., 2021). For instance, potential seasonal variation in LULC estimates should considered during an LULC update in a modeling approach such as Hales et al. (2023), where a global hydrologic model GEOGloWS is bias corrected for extreme event forecasting in underdeveloped regions using a single instance of Dynamic World data. Our findings show that there is the potential for discrepancies at least for temperate watersheds in the eastern United States if the season of LULC update were not accounted for. These illogical LULC changes could also be pertinent for models that can use a mosaic*

approach to represent spatial variability of LULC within coarser grid cells (e.g., CLM5; Lawrence et al., 2019). The mosaic approach assumes that land surface properties (e.g., water fluxes) are homogeneously related to the LULC type (Li et al., 2013; Qin et al., 2023), in which case an illogical conversion of 12% area from forest to other types (our Case #3 example) could carry forward into the models, and potentially impact water and energy flux estimates or parameterizations similar to an actual LULC change. For instance, deforestation has previously been shown to alter heat and carbon fluxes and ecosystem productivity in CLM5 (Marufah et al., 2021; Luo et al., 2023). Variability within input data sub-grids has also been shown to influence model parameter optimization and performance simulating hydrology, making it an important aspect to account for (Samaniego et al., 2010)." (page 16, line 346 to page 17, line 361)

Page 4, lines 100-101. This phrasing implies more performance measures were used. Please list these here or change the text accordingly.

Response: We reworded to be clearer.

"*Performance measures R^2 and root mean square error (RMSE; Willmott et al., 1985) were used to compare models from different seasons.*" (page 4, lines 116-117)

Page 4, line 101. I don't think this is an appropriate reference to RMSE

Response: We changed the RMSE reference to Willmott et al., 1985.

Figure 2. It would be helpful to list the models more specifically. E.g., "regression models", "SWAT", "SWAT" I believe

Response: We made the improvement as suggested.

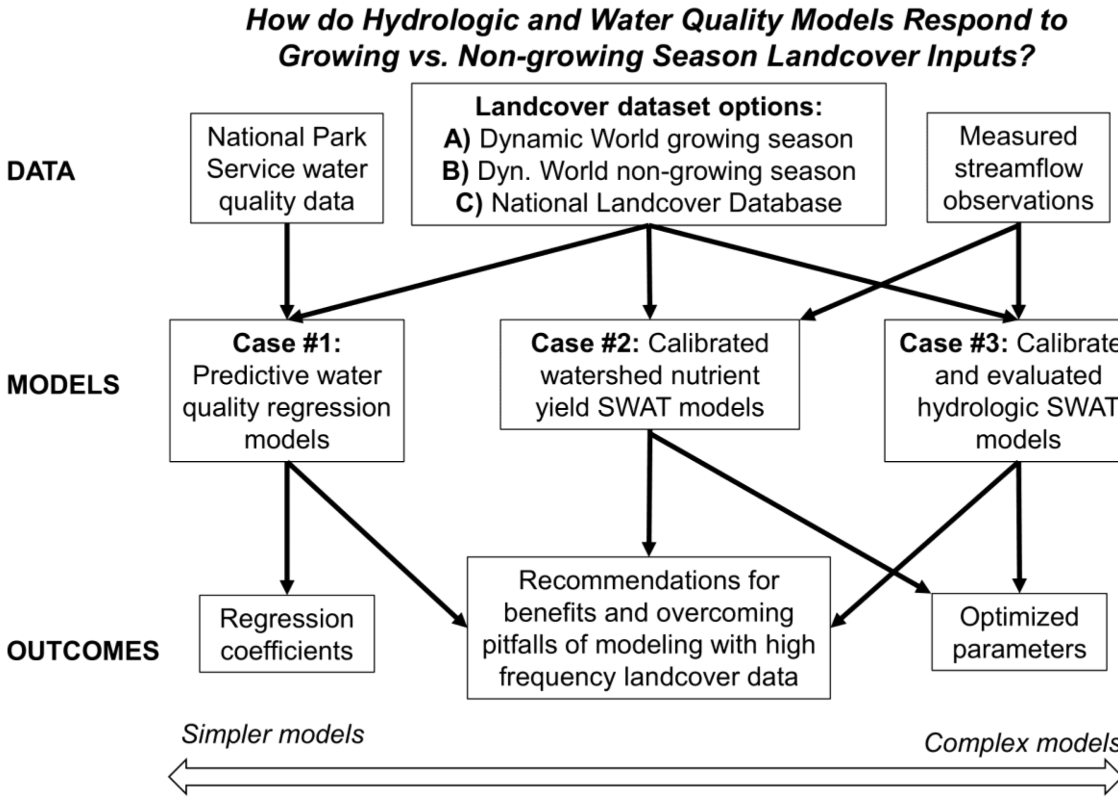

Figure 2: Conceptual diagram of the study.

Figure 2. This figure seems ever so slightly blurry to me, and this is a bit worse in most of the following figures. I'd suggest exporting the figures at 300 dpi to avoid excessive blurring.

> Response: We have increased the resolution of the figures in the manuscript as suggested. For production we would upload vector (or in some cases very high dpi) images.

Page 6, lines 126-127. Were these HRUs recreated with the growing/non-growing LULC maps?

> Response: Yes, switching the LULC layer resulted in recreating HRUs. We updated the text to include this detail.
>
> "*We created a new SWAT LULC look-up table for QSWAT to read Dynamic World data and recreate HRUs.*" (page 6, lines 148-149)

Page 6, lines 127-128. I think this paper needs a dedicated table that shows how HRUs are defines for each of the test cases. This table should list inputs into the tool, as well as the resulting number of HRUs in each test case.

Figures showing for each study site how many HRUs were defined and what the defining characteristics of the individual HRUs are would be even better.

> Response: We have now added two figures that detail how HRUs are being populated with Dynamic World 2016 (growing vs. non-growing season) and NLCD 2016 LULC for each case (Figures S4 and S5; included in the response to comment "2 Materials and Methods. SUMMARY" above). Figure captions include exact HRU numbers for reference. Additionally, we now describe HRU workings in our results for each case.
>
> Case #2: "*The differences observed between models using Dynamic World LULC were due to the 9% increase in built areas in non-growing season Dynamic World 2016 data, which have more impervious surfaces, a higher runoff curve number, and generate proportionally more water and nutrient runoff than the forested areas which were classified during the growing season. This could be particularly problematic when using computationally more efficient SWAT models that assign subbasin conditions based on the dominant HRU, as a change in dominant LULC type in a watershed could result in different subbasin conditions in the model greater than the proportional change in LULC. In this case, using non-growing season instead of growing season LULC input caused the model to switch two HRUs representing 21.9% of watershed area from being populated with Dynamic World trees LULC class to built LULC class (Figures S4 and S5). For perspective, the nutrient outputs for the SWAT model with Dynamic World 2016 growing season LULC were similar to those simulated by the SWAT model with NLCD 2016 LULC input using the same parameter adjustments (Figure 5c).*" (page 13, lines 286-295)
>
> Case #3: "*At the HRU level, using growing vs. non-growing season Dynamic World 2016 LULC in this case resulted in a 12.8% change in model HRU trees proportions, which is proportionate to the change in input trees estimates, as would be expected with the maximum HRU designation approach (Figures S4 and S5).*" (page 14, lines 309-312)
>
> We also added discussion of the implications of different approaches to populate HRUs, in response to your comment beginning "Page 9, lines 204-209" below. In summary, recommending maximizing HRUs as much as possible and evaluating the new approach you introduced there.

Page 6, lines 139-140. Is there some evidence to support that 2500 iterations is a sufficient number? Intuitively this feels low to me.

> Response: We followed a previous calibration using this approach in Nguyen et al. (2022), which used 1,000 iterations. Nguyen's was based on the approach developed in Abbaspour et al., (2004), which recommended approximately 1,000 to 2,000 iterations.

We chose to use more iterations than these references since our model was computationally efficient. We added this additional support to the text.

"*The parameters optimized during the Latin hypercube approach, which had 2,500 iterations (based on Nguyen et al., 2022 and Abbaspour et al., 2004), are shown in Table S3.*" (page 6, lines 163-165)

Page 6, line 141. Was the model run at a monthly time step too, or at a higher temporal resolution?

Response: Yes, this case was at the monthly time step. We updated the text to be clearer here.

"*Calibration and evaluation data were complete monthly streamflow (n=108 months) and nitrogen (n=10 months) data from the USGS station 01648010 (concentrations converted to loads by multiplying by streamflow), split with the first half for calibration and the latter half for evaluation at the monthly time step.*" (page 6, line 165 to page 7, line 167)

Page 6, lines 149-151. Given this information, Latin-Hypercube-based calibration with only 2500 iterations seems really low to me. Can the authors clarify this discrepancy in sampling numbers and why these numbers (2500 LHS iterations for calibration; 8500 samples for SA; 3200 iterations for post-SA calibration) are appropriate for each case?

Response: We explain the Latin hypercube samples two comments above. For the PAWN sensitivity analysis and AMALGAM calibration, we based our iterations after previous work, which used 6,000 samples for PAWN sensitivity analysis and 3,200 AMALGAM calibration iterations (Myers et al., 2021a). The number of PAWN sensitivity analysis iterations for this work is larger than Myers et al. (2021a) because we were able to fit the extra iterations of our model within computing restraints, while the number of calibration iterations was the same. In Myers et al. (2021a), a detailed exercise was performed to test for equifinality using these PAWN and AMALGAM iteration numbers, which identified that the variation in optimized water balances due to equifinality of parameters was relatively much smaller than that caused by choosing different calibration/validation time periods, supporting that the number of iterations was sufficient to find global optima in their cases. We have updated the text with these additional details.

"*Eight thousand SWAT model runs with growing season Dynamic World 2016 data were used for the sensitivity analysis, based on Myers et al. (2021a).*" (page 7, lines 174-175)

"*We then calibrated the Difficult Run Watershed SWAT models at the daily time step using the AMALGAM optimization algorithm (Vrugt and Robinson, 2007) with 3200 iterations (based on Myers et al., 2021a)...*" (page 7, lines 176-178)

Table 1, CH_KII.rte. Is this a realistic change in parameter values? The documentation (https://swat.tamu.edu/media/69374/ch25_input_rte.pdf) seems to say this parameter represents near-stream soil conductivity. This conductivity going op by a factor 35 as a result of seasonal land cover change seems unrealistic to me. Can this be explained?

<u>Response</u>: No, it is not an actual change in CH_KII, rather it is an outcome of using different seasons of LULC input in the models (what we aimed to test). As it would not be feasible to know real values of parameters such as CH_KII in the watersheds of this study (i.e., an intense field campaign), we inversely obtain them through calibration within reasonable parameter ranges. The calibration of CH_KII is commonly done using absolute values of mm/hr as we did, following this approach (e.g., Nguyen et al., 2022; Myers et al., 2021a).

Our findings of different parameter optimizations depending on the season of LULC estimates inputted into the model demonstrate the impacts that illogical LULC changes can have on this process, leading to changes in the optimized parameter values such as these. We find this to be a major take-away and fitting with the purpose for our analyses.

"*For the independently calibrated models (Case #3), we assessed the performance of seasonally tuned models rather than the single model of the LULC change case, to provide fairer comparison of calibrated model performances since each model was optimized to its unique LULC situation.*" (page 5, lines 124-126)

We further describe the implications of this sensitivity of parameter optimizations to seasonal LULC data for the fine-tuning of hydrologic and water quality models.

"*It is critical to consider that the differences in parameter values create the potential for the models to respond differently to future changes in LULC or climate change due to variations in unmeasured water balance outputs (Myers et al., 2021a).*" (page 16, lines 331-333)

"*Environmental and geospatial researchers should be aware of this sensitivity when developing models and assessing changes in LULC as they relate to water quantity and quality, especially when considering the use of different seasons of available LULC data in a model.*" (page 18, lines 404-406)

Table 1, SOL_BD.sol. According to the docs "The soil bulk density expresses the ratio of the mass of solid particles to the total volume of the soil" (). Is it realistic to see this change by almost 20% as a function of growing/non-growing seasons?

I would be surprised if the soil properties change to this extent on a seasonal cycle. It seems more likely to me that the change in land-use type leads to a different set of soil properties, but this would be a consequence of SWAT only accepting the dominant land type - and not of a natural process that the model replicates.

> Response: Similar to our above response, this is not an actual change in soil bulk density, but rather a potential implication of using different seasons of LULC input in a hydrologic model. Our calibration approach allows the regional soil bulk density parameter to vary relatively by ±20% when optimizing hydrologic processes for streamflow simulation (in other words, we assume our input values could be off by ±20%). We found that there can be different optimizations of soil bulk density within this window when a model is developed using an LULC layer from growing vs. non-growing season. Also, the model in this case was developed so that the maximum number of HRUs can be incorporated to compare model performance between the independent cases. This is pertinent because it provides guidance about the potential sensitivities of using LULC estimates from different seasons in hydrologic and water quality models.

Table 1, SNOCOVMX.bsn. How can the snow depth for full vegetation cover be higher in the non-growing season than in the growing season? Wouldn't it make more sense to be the other way around, with lower vegetation height in the dormant season?

More generally, this suggests the calibration approach deserves further attention. There may be:
- equifinality in optimal parameter values (i.e., multiple different parameter sets give the same nominal performance scores). This may need to be accounted for through looking at multiple calibrated parameter sets.
- compensation of any number of data or model errors through unrealistic parameter values.

> Response: Similar to the above two responses, we highlight that the models in this case are calibrated independently of one-another. Parameter values optimized for a model developed with the growing season LULC input could be different from those for a model developed with the non-growing season LULC input, as we witness with the different SNOCOVMX optimizations, that do not align with an actual change in the watershed. The finding you describe is one of the key takeaways of our work: different seasons of LULC input could lead to different optimizations of parameters, as has been previously found with actual LULC change, and we discuss that this has implications for using a model to simulate actual changes such as LULC or climate.

"*The difference in forests of 12% of watershed area between growing and non-growing season Dynamic World 2016 data for Difficult Run is as large a difference as real changes in forests that have been found to cause these sensitivities in model parameters (Li et al., 2019), but was likely caused by classification variation rather than an actual cycle from trees to built area and back (Hermosilla et al., 2018).*" (page 16, lines 328-331)

As mentioned above, this can have implications for simulating changes in the watershed that we want to ensure modelers are aware of.

"*When seasonal changes in LULC data occur, due to classification difficulties such as vegetation cycles (e.g., deciduous leaf cover in mixed-LULC areas), hydrologic and water quality models developed using growing season LULC inputs could behave differently from those using non-growing season LULC, with meaningful differences for environmental efforts such as pollutant load reduction targets.*" (page 18, lines 399-402)

We also added a description about the potential for equifinality in our work, based on previous studies that used large numbers of model ensembles and different parameter set initializations in the optimization algorithm.

"*Also, although we did not investigate equifinality using model ensembles for this case, we aimed to limit it by employing a calibration approach with multiple optimization algorithms (AMALGAM; Vrugt and Robinson, 2007) and calibrating only the most sensitive parameters. Previous work has found this approach to be robust to equifinality relative to other factors affecting parameter optimizations such as calibration/validation time period selection (Myers et al., 2021a) and model structures (Myers et al., 2021b), and our findings are in line with previous investigations of LULC input changes impacting SWAT model parameter optimizations (such as forest conversion causing runoff curve number adjustment to vary relatively by 21%; Li et al., 2019).*" (page 16, lines 333-339)

We hope this clarifies the purpose and implications of this case. We appreciate the thorough work of the reviewer to ensure our conclusions are reliable and make sense.

3.1. Seasonal landcover comparisons. SUMMARY:
I believe this subsection needs some dedicated text explaining the realism of the identified seasonal LULC changes.

<u>Response</u>: We here discuss the realism of the seasonal LULC changes, noting that they are unlikely to be actual changes but rather illogical classification inconsistencies.

"*Actual on-the-ground changes from built LULC to other types, or from other LULC types to trees (e.g., succession), are not likely to be occurring within the short (seasonal) time interval between our LULC composites (Cai et al., 2014).*" (page 10, lines 231-233)

"*Potential causes for these differences include vegetation phenology (e.g., green up) affected by climate (Khodaee et al., 2022), or measurement artifacts such as atmospheric conditions (aerosol scattering, water vapor, and absorption of light) and reflectance (bidirectional reflectance and zenith angle) which can cause non-random errors in top-of-atmosphere readings used for classifying LULC (Zhang et al., 2018; Kaufman, 1984; Rumora et al., 2020). Dynamic World used a calibrated surface reflectance product to train the classifier (Sentinel-2 Level-2A; L2A) but a top-of-atmosphere product (Sentinel-2 Level 1C; L1C) to generate the dataset (Brown et al., 2022). Previous work in our study area has found strong inter-annual variations across spectral bands in remotely sensed imagery that were caused by uncorrected atmospheric conditions and could impact multi-year LULC classification (Sexton et al., 2013). These differences in atmospheric conditions and reflectance would not be corrected for in Dynamic World data and potentially contribute to differences in classification results over time.*" (page 10, lines 218-227)

Figure 3. Is there some sort of physical evidence that supports these rather large seasonal changes in built-up and tree cover?

If there isn't, that implies that the growing/non-growing LULC maps are not particularly accurate, and that turns this manuscript into more of an academic/hypothetical exercise (i.e., "what are the consequence if LULC were to change substantially on a seasonal basis") rather than something with immediate practical relevance.

Response: No, there is not physical evidence, and we describe that the seasonal LULC changes between built area and trees identified in our investigation are illogical changes (see response above). We strongly believe these temporal inconsistencies are of practical relevance to modelers using high spatiotemporal resolution LULC classifications such as Dynamic World. As we demonstrate, models can simulate illogical LULC changes as a real change, impacting simulation results in ways that do not reflect actual LULC transitions. Thus, modelers should take care and understand the potential impacts of these changes when developing studies and interpreting results of the high spatiotemporal resolution LULC data. The relevance of our findings becomes even more immediate considering that hydrologic and water quality models are frequently used for decision-making purposes with real-world implications, making these seasonal LULC changes essential to our aim to provide guidance to future modelers.

For instance, recent work used one instance of Dynamic World land cover data to bias correct a global hydrologic model to be used for forecasting hydrologic extremes and potential emergencies in underdeveloped countries (Hales et al., 2023). If that correction were used with a different instance of LULC data from a different season, our findings show that there could potentially be illogical LULC discrepancies for the temperate eastern United States watersheds. This is particularly important when models are being used for such meaningful societal purposes.

Incorporating potentially illogical land cover changes into modeling inputs could have implications to HESS readers, particularly when it is not clear which seasonal changes are real and should have logical effects on model outputs, or which are merely illogical errors that should be corrected before influencing model outputs.

We explain the nature of these LULC changes in our abstract and also our results and discussion (noted in the response above):

"*Non-growing season data resulted in LULC classifications that had more built area and less tree cover than growing season data due to seasonal impacts on classifications rather than actual LULC changes (e.g., quick construction or succession). In mixed-LULC watersheds, seasonal LULC classification inconsistencies could lead to differences in model outputs depending on the LULC season used, such as differences in watershed nitrogen yields simulated by the Soil and Water Assessment Tool.*" (page 1, lines 20-24)

Page 9, lines 200-203. I think this is important in relation to my previous comment. If there are doubts about the accuracy of the classifications, the type of question being asked in this manuscript changes to a "what if" kind of investigation. Note that this is not necessarily worse than a "this happens in reality, here are the implications for modelling" kind of study, but this will need to be clarified in the framing of the work.

Response: Our aim was to investigate how hydrologic and water quality models respond to growing season vs. non-growing season classified LULC, to inform future geospatial modeling efforts using high spatiotemporal resolution LULC data such as Dynamic World. We found that using a different season of LULC can potentially impact model outputs or parameterizations due to illogical LULC changes. As we describe above, these are situations that can be encountered by modelers using the data, and we provide guidelines for how to deal with the situations. To clarify the framing of our work, we updated our communication of objectives in the introduction to make it clearer that these illogical LULC changes could be encountered by modelers as well.

"*We used the Dynamic World LULC dataset to demonstrate how estimates of LULC can change between the growing and non-growing seasons (note that estimates of LULC could change due to real transitions or due to illogical classification inconsistencies described above).*" (page 2, lines 64-67)

We also note this in our experiment design.

"*For the LULC change simulation (Case #2), we evaluated how a model calibrated to one LULC season could respond to LULC data from another season, such as when simulating impacts of a watershed LULC change, particularly with regards to sensitivity to potential illogical LULC transitions in the high temporal frequency data.*" (page 4, line 117 to page 5, line 119)

Page 9, lines 204-209. It's not clear to me if this has been done in this paper, but one way to address this from a modelling angle would to ensure the HRUs are defined based on seasonal stability of dominant LULC:
- HRU 1: classified as forest in both growing and non-growing season
- HRU 2: classified as [something else] in both growing and non-growing season
- HRU3: classified as forest in growing and [something else] in non-growing season.

Such an approach would isolate the impact of the seasonal changes in LULC to the specific areas where there is in fact a change, and (presumably) show a reduced impact of this problem in modelling outcomes (because most of the model setup would -correctly- remain unaffected by the seasonal LULC change).

If this is in fact what the authors did I would strongly suggest to clarify the HRU delineation procedure, because this was not clear to me.

Response: This approach to delineating HRUs would be an interesting avenue for future research building on our work. We explored traditional approaches in this study (dominant subbasin HRU for Case #2, and maximum HRU resolution for Case #3). We appreciate the positive direction for brainstorming how to improve modeling procedures to account for the seasonal LULC classification inconsistencies, and think evaluation of this idea could be impactful to other modelers using the high spatiotemporal frequency data. Thus, we added a discussion of this topic and how it fits with our results.

"*When using seasonal LULC estimates in hydrologic and water quality models, we recommend differentiating HRUs as much as possible (like our maximum HRU resolution approach for Case #3) so that the potential for disproportionate impacts from LULC season is minimized. Aggregating HRUs by dominant characteristics over an area may lead to high variability in responses depending on areas where estimated LULC changes are substantial enough to switch dominant HRU LULC characteristics, which in our second case was two HRUs in the northern part of the watershed. However, future work could investigate approaches to differentiate HRUs that further limit or remove the impacts of seasonal variation in LULC estimates, such as separating areas with stable LULC across seasons from those with substantial LULC variability, to isolate the most affected parts of the watershed. Thus, HRUs that remain unaffected by seasonal changes* *in LULC estimates would be preserved, while HRUs with potential for change due to illogical seasonal LULC transitions could be identified and treated separately. In this proposed approach, aggregating HRUs may be possible to resist disproportionate impacts of LULC seasonality while alleviating computational burdens of large HRU numbers. Evaluation of such an approach could help advance the hydrologic and water quality modeling community into higher spatiotemporal resolution LULC capabilities.*" (page 17, lines 368-379)

3.2 Case #1: Water quality regressions. SUMMARY:
No real comments on this subsection.

3.3 Case #2: Hydrologic and nitrogen yield models. SUMMARY:
The text in this subsection seems factually correct to me, but I find it difficult to interpret and judge its relevance. This is mainly due to several things that are unclear to me about how the models were configured, and to what extent these findings are novel.

I would strongly recommend that the authors further clarify their experimental design and add a section to their introduction that explains the current state of understanding within the SWAT community about this topic.

> Response: We have updated the introduction and discussion to further highlight existing understanding within the SWAT community, detailed in our responses to the comments beginning with "Page 11, lines 238-243" and "1 Introduction. SUMMARY." Beyond these responses, we also add more details about SWAT sensitivities to our introduction here:
>
> "*These models are known to be sensitive to actual LULC changes over longer (e.g., 10+ year) time spans, such as forests being converted to other LULC types (Li et al., 2019; Basu et al., 2022).*" (page 2, lines 46-48)
>
> We also direct you to the comment beginning with "Page 6, lines 127-128" above, where we clarify our experimental design with regards to HRU population and provide additional updates to the text to help interpret results.

Page 11, lines 229-231. Are these scores indicative of good performance? In other words, is this within expectations for SWAT in general and this watershed specifically? Some references to back this up would be good.

> Response: We now add references to put our evaluation results into perspective with other similar studies.

"*As these values are similar to those of previous SWAT evaluations in urban watersheds that occurred at monthly time steps (Basu et al., 2022; Halefom et al., 2017) and other work with multiple calibration variables (e.g., Myers et al., 2021b), we concluded that the model developed with Dynamic World 2016 growing season data was reliably simulating real conditions at the monthly time step.*" (page 11, line 255 to page 12, line 259)

Page 11, lines 231-233. It's unclear to me if the growing season LULC was used for the whole simulation (including the non-growing) season months or if the LULC changes halfway through the simulation. I suspect it's the former based on Table 3. Why is it theoretically sound to use dedicated growing season LULC for a year-round simulation, and vice versa?

Response: The theoretical soundness of this is related to what our study investigates: traditionally, watershed modelers will use an LULC dataset classified primarily in the growing season to simulate hydrology and water quality year-round (e.g., NLCD in (Botero-Acosta et al., 2022; Avellaneda et al., 2020). The release of the analysis-ready global Dynamic World dataset has facilitated opportunities for modelers to use LULC classified at different times of the year, such as the non-growing season only, with high (10 m) spatial resolution. For instance, there could be opportunities to model LULC impacts to water quality representing high temporal resolution changes or near-current conditions. We aimed to evaluate the impacts of inputting different seasons of LULC data in models to inform the modeling community as we advance into these higher spatiotemporal resolution frameworks for watershed studies. In each case, the designated LULC input (growing or non-growing season) was used for the full simulation to evaluate differences over the same time period. Thus, we could compare models objectively with the simulation period (and other characteristics such as weather) controlled. We updated our text here to make this clearer.

"*When the calibrated parameter adjustments were transferred to the SWAT model developed with non-growing season LULC (as could be done when simulating an actual LULC change), streamflow performance decreased by approximately 0.30 NSE units and nitrogen yield PBIAS became -34.4% to -57.4%, implying overestimation of nitrogen. Note that both models were run over the same time period to compare performance.*" (page 12, lines 259-262)

Page 11, lines 238-243. This may be a (very) relevant finding for practical use, but it is not very clear to me to what extent this is already known. Theoretically at least I don't find it very surprising that a model calibrated for conditions X does not necessarily do well under substantially changed conditions Y. Can the authors add some discussion to highlight to what extent this finding is known or new within the SWAT community?

Response: We agree with the reviewer of the relevance and practicality of these findings. As suggested, we now expand our discussion here about how the novel findings fit into SWAT literature specifically.

"*This aligns with previous work that found impacts of actual LULC changes on hydrologic model performance, albeit at longer (e.g., 10+ year) time spans (Li et al., 2019). Although hydrologic and water quality models such as SWAT are often developed using LULC classified primarily in the growing season (e.g., Botero-Acosta et al., 2022; Avellaneda et al., 2020), the availability of analysis-ready seasonal LULC data such as Dynamic World makes evaluations of LULC estimate sensitivity at shorter (i.e., seasonal) time spans pertinent.*" (page 12, lines 270-274)

For background, our primary goal here was to take a traditional calibration approach to growing season data and evaluate what would happen when non-growing season data is used, to provide guidance for using the high spatiotemporal frequency LULC data. We are not aware of any previous SWAT literature that investigates this. However, analysis-ready data such as Dynamic World are at the forefront of non-growing season LULC data being built into modeling approaches, which we discuss with our future directions. We also introduce that hydrologic and water quality models such as SWAT are traditionally developed with LULC classified primarily during the growing season.

"*Studies relating hydrology and water quality to LULC often use an LULC dataset developed primarily from growing season data, such as the United States National Landcover Database (NLCD; Jin et al., 2019) or Cropland Data Layer (CDL; Boryan et al., 2011), and/or use an LULC dataset available at an annual time step (Sulla-Menashe and Friedl, 2018; Buchhorn et al., 2020; Gray et al., 2022).*" (page 1, lines 31-34)

"*As models advance into higher spatiotemporal resolution following increasing computational resources and data availability (e.g., Hales et al., 2023), we encourage the modeling community to be cognizant of the potential impacts of illogical seasonal LULC change, such as we identified for mixed LULC areas of the eastern United States.*" (page 17, lines 361-364)

We thus believe that these findings are novel and pertinent not only to SWAT literature, but also to the broader hydrologic and water quality modeling community. Particularly with regards to our evaluations of inputting Dynamic World into models, which could be done with other models beyond SWAT and regressions. We note that our preprint has been cited in two independent peer-reviewed studies advancing knowledge of LULC change impacts to freshwater systems, which may be examples of this interest (Dede et al., 2024; Giuliani, 2024).

Table 3, calibration. It would be good to clarify somehow in this table that this row and the one below refer to transferred parameter performance, and not a new calibration exercise.

> Response: We made the improvement as suggested and added the clarification to the caption.
>
> "*Table 3: Model performance metrics for the calibrated Rock Creek hydrologic model (Case #2) for streamflow and nitrogen yield, based on Nash Sutcliffe Efficiency (NSE), mean absolute error (MAE), and percent bias (PBIAS, where <0 implies overestimation bias), at the monthly time step. In this case, model parameters were all calibrated to growing season Dynamic World 2016 data to investigate the impacts of simulating an LULC change using non-growing season data (e.g., the optimized parameter adjustments were kept the same).*" (page 12, lines 276-279)

3.4 Case #3: Independently calibrated hydrologic models. SUMMARY:
I think this subsection needs some extra discussion about how good these simulations actually are, in spite of the scores reported, and what it means that the differences between the various simulations are much smaller than the differences between simulations and observations.

> Response: We have expanded on the discussion of model performance as suggested. Also, we note that similar to Basu et al. (2022), our objective was to examine differences due to LULC holding all other model data inputs the same between simulations.
>
> "*These are in line with satisfactory performance from previous work, particularly considering the daily time step (Moriasi et al., 2007; Kalin et al., 2010; Basu et al., 2022).*" (page 13, lines 303-304)
>
> "*Discrepancies such as underestimated low flows or peaks could reflect difficulties simulating hydrology in urban areas with complex stormwater pathways, as the Difficult Run Watershed was 58% developed area in the NLCD 2016 data. Also, differences between independently calibrated streamflows could be smaller than differences with observed data, which could be due to uncertainties in other non-LULC model inputs shared among the calibrations (Basu et al., 2022).*" (page 14, lines 306-309)

Page 12, lines 269-270. Two interesting features I noticed here are that:
1. Both growing (red) and NLCD (yellow) show a serious underestimation bias for low discharges, but the non-growing simulations seem to do a little better at the low flows.
2. Generally speaking, the simulated hydrographs in 6d really don't seem that good to me. The lows are underestimated, mediums overestimated (Mar-Apr 2016), peaks are underestimated or missed completely (Sep 2016). The differences between the different simulations seem much smaller than the model errors in general - what sort of conclusion should we draw from this?

Response: This builds on our response to the comment above. The main conclusion is that the independently calibrated models all simulated streamflow with comparable performance relative to one another, which was what we aimed to evaluate in our objectives.

Since we were interested primarily in LULC change impacts, we standardized all other input data among the models (e.g., soils, weather, and topography) so that they can be more directly comparable. Thus, errors related to uncertainties in those other inputs could carry forward in similar ways for each model. Also, complex urban stormwater pathways and stormwater infrastructure varying between HRUs could have greater impacts to model responses compared with observed data, than to model responses relative to the other models. Using the daily time step, our models in this case undergo much more temporal scrutiny compared to studies that use longer (e.g., monthly) time steps in heavily urbanized watersheds such as Basu et al. (2022) and Halefom et al. (2017).

Despite these modeling difficulties, we found that in these urban/mixed watersheds, the seasonal LULC classification differences between built area and trees can be most prevalent, which makes it important to compare relative model performance here. Thus, we believe our models are sufficient to demonstrate the impacts of the seasonal LULC classification differences because they behave similarly to other SWAT evaluations in heavily anthropogenically altered watersheds, align with our objectives, and can generally be considered sufficient based on performance statistics (Moriasi et al., 2007; Kalin et al., 2010). We also find it important we communicate these evaluations clearly, including the features you mention, which we incorporated into our response to the comment above.

3.5 Future directions. SUMMARY:
The main conclusion in this subsection seems to be that using seasonal LULC changes can be to some (a large?) extent consequences of misclassification, and that this impacts model simulations by providing the model with unrealistic inputs.

I don't disagree with this but it's largely unclear to me what the novelty of these findings is. There is some documentation available that already claims the LULC changes on annual satellite products should be treated carefully due to classification issues (e.g. MODIS), and it should come as no surprise that giving a model incorrect inputs is going to lead to unhelpful outputs ("garbage in, garbage out"). I would strongly recommend that the authors clarify the novelty or relevance of these findings fort the wider HESS audience.

For what it's worth, in my experience a typical way of dealing with LULC classification uncertainty is to get multiple maps at different points in time, and either use or assess changes in the median or majority land cover identified on those maps. This reduces the classification uncertainty by generating multiple samples of the same area of interest, and looking for common patterns less affected by variability at specific points in time.

Response: Evaluation of LULC products at high spatiotemporal resolution has been described as an important research need with vast societal implications (Radeloff et al., 2024). Our study is one step toward achieving this goal, providing guidance to the modeling community for using data such as Dynamic World. In addition to our general contributions, as far as we know, we are the first study to evaluate the temporal consistency of Dynamic World and modeling implications. We also are not aware of other studies that have critically evaluated these data in a hydrologic and water quality modeling framework such as SWAT. Thus we state:

"*This high spatiotemporal resolution creates unprecedented opportunities for modelers to study the impacts of phenomena such as emerging settlements, agricultural dynamics, and forest conversion on outputs such as ecosystem dynamics and biogeochemical budgets (Brown et al., 2022). For environmental research to take advantage of these high temporal resolution data, we need to understand the impacts of potential seasonal variation in LULC estimates on geospatial models, which use LULC data to support water resources management across the globe (Fu et al., 2019; Guo et al., 2020; Murphy, 2020).*" (page 2, lines 54-59)

Evaluations such as we provide are critical for providing the modeling community with guidance for using sub-annual, high spatiotemporal resolution LULC data in their simulations. For instance, in heterogeneous areas where trees and built area are mixed, the seasonal consistency of data can be important for models of LULC change impacts to hydrology, even among LULC instances which might not appear unrealistic or incorrect on their own (Brown et al., 2022 demonstrated that individual images of Dynamic World LULC evaluate well against human-classified reference). Thus, we do not believe that our work is a case of "garbage in, garbage out". Rather, it is a critical evaluation to inform the design of future work and interpretation of modeling results, particularly when models would be used in decision making with real-world implications for freshwater management.

We have now provided more details on the novelties of our work to make our position within existing literature clearer, such as the additions noted in our responses to the above comments beginning with "1 Introduction. SUMMARY," "Page 9, lines 204-209," "3.3 Case #2: Hydrologic and nitrogen yield models. SUMMARY," and "Page 11, lines 238-243."

Also, our design used the majority (mode) classification from combining individual instances of LULC over the growing vs. non-growing seasons, as recommended by Brown et al. (2022).

"*We used Google Earth Engine (Gorelick et al., 2017) to generate a different Dynamic World LULC dataset for growing season (spring equinox to autumn equinox, 2016) and non-growing season (autumn equinox, 2015 to spring equinox, 2016) for the monitored watersheds by taking dominant LULC for each pixel over these time periods, following the suggested approach (Brown et al., 2022).*" (page 4, lines 95-98)

4 Conclusions. SUMMARY:
The same comment I left at Section 3.5 applies here. I agree with the stated conclusions but it is unclear to me to what extent these are sufficiently novel and/or relevant to warrant publication in HESS.

Listing examples where this issue actually occurs (i.e. studies using growing season LULC estimates in the way investigated in this manuscript) would at least partly address this relevance issue, though my question about novelty would remain.

Response: We hope our explanations and additions to the text in the response above (as well as the additional locations mentioned there) will communicate better the novelties and relevance of the work. In summary, we provided references to other studies using LULC classified primarily in the growing season. We then expanded on our novelties evaluating how sub-annual LULC inconsistencies in high spatiotemporal resolution data can affect hydrologic and water quality models, and provided guidance for modelers using data such as these. Our work now provides additional insights on data selection for LULC change simulation, objective model performance comparisons, HRU development, and impacts on parameterizations for hydrologic and water quality modelers using high spatiotemporal frequency LULC data such as the Dynamic World product, beyond what can be found elsewhere on the topic. We illustrate these using cases of models ranging from simple to more complex. We expect these novelties will be most relevant when modelers are designing studies and interpreting results with real-world decision making implications for freshwater management.

We again acknowledge you Dr. Knoben for your help reviewing and improving the quality of our manuscript, and are happy to answer any additional questions.

References:

Abatzoglou, J. T.: Development of gridded surface meteorological data for ecological applications and modelling, International Journal of Climatology, 33, https://doi.org/10.1002/joc.3413, 2013.
Abbaspour, K. C., Johnson, C. A., and van Genuchten, M. Th.: Estimating Uncertain Flow and Transport Parameters Using a Sequential Uncertainty Fitting Procedure, Vadose Zone Journal, 3, 1340–1352, https://doi.org/10.2136/vzj2004.1340, 2004.

Arabi, M., Frankenberger, J. R., Engel, B. A., and Arnold, J. G.: Representation of agricultural conservation practices with SWAT, Hydrol Process, 22, https://doi.org/10.1002/hyp.6890, 2008.

Avellaneda, P. M., Ficklin, D. L., Lowry, C. S., Knouft, J. H., and Hall, D. M.: Improving Hydrological Models With the Assimilation of Crowdsourced Data, Water Resour Res, 56, e2019WR026325, https://doi.org/10.1029/2019WR026325, 2020.

Basu, A. S., Gill, L. W., Pilla, F., and Basu, B.: Assessment of Variations in Runoff Due to Landcover Changes Using the SWAT Model in an Urban River in Dublin, Ireland, Sustainability 2022, Vol. 14, Page 534, 14, 534, https://doi.org/10.3390/SU14010534, 2022.

Boryan, C., Yang, Z., Mueller, R., and Craig, M.: Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program, http://dx.doi.org/10.1080/10106049.2011.562309, 26, 341–358, https://doi.org/10.1080/10106049.2011.562309, 2011.

Botero-Acosta, A., Ficklin, D. L., Ehsani, N., and Knouft, J. H.: Climate induced changes in streamflow and water temperature in basins across the Atlantic Coast of the United States: An opportunity for nature-based regional management, J Hydrol Reg Stud, 44, 101202, 2022.

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., and Tait, A. M.: Dynamic World, Near real-time global 10 m land use land cover mapping, Scientific Data 2022 9:1, 9, 1–17, https://doi.org/10.1038/s41597-022-01307-4, 2022.

Buchhorn, M., Lesiv, M., Tsendbazar, N. E., Herold, M., Bertels, L., and Smets, B.: Copernicus Global Land Cover Layers—Collection 2, Remote Sensing 2020, Vol. 12, Page 1044, 12, 1044, https://doi.org/10.3390/RS12061044, 2020.

Cai, S., Liu, D., Sulla-Menashe, D., and Friedl, M. A.: Enhancing MODIS land cover product with a spatial-temporal modeling algorithm, Remote Sens Environ, 147, https://doi.org/10.1016/j.rse.2014.03.012, 2014.

Dede, M., Sunardi, S., Lam, K. C., Withaningsih, S., Hendarmawan, H., and Husodo, T.: Landscape dynamics and its related factors in the Citarum River Basin: a comparison of three algorithms with multivariate analysis, Geocarto Int, 39, https://doi.org/10.1080/10106049.2024.2329665, 2024.

Dow, C. L. and Zampella, R. A.: Specific Conductance and pH as Indicators of Watershed Disturbance in Streams of the New Jersey Pinelands, USA, Environmental Management 2000 26:4, 26, 437–445, https://doi.org/10.1007/S002670010101, 2000.

Fu, B., Merritt, W. S., Croke, B. F. W., Weber, T. R., and Jakeman, A. J.: A review of catchment-scale water quality and erosion models and a synthesis of future prospects, Environmental Modelling & Software, 114, 75–97, https://doi.org/10.1016/J.ENVSOFT.2018.12.008, 2019.

Fuka, D. R., Easton, Z. M., Brooks, E. S., Boll, J., Steenhuis, T. S., and Walter, M. T.: A Simple Process-Based Snowmelt Routine to Model Spatially Distributed Snow Depth and

Snowmelt in the SWAT Model, J Am Water Resour Assoc, 48, 1151–1161, https://doi.org/10.1111/j.1752-1688.2012.00680.x, 2012.

Giuliani, G.: Time-first approach for land cover mapping using big Earth observation data time-series in a data cube–a case study from the Lake Geneva region (Switzerland), Big Earth Data, https://doi.org/10.1080/20964471.2024.2323241, 2024.

Glotfelty, T., Ramírez-Mejía, D., Bowden, J., Ghilardi, A., and West, J. J.: Limitations of WRF land surface models for simulating land use and land cover change in Sub-Saharan Africa and development of an improved model (CLM-AF v. 1.0), Geosci Model Dev, 14, https://doi.org/10.5194/gmd-14-3215-2021, 2021.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sens Environ, 202, 18–27, https://doi.org/10.1016/J.RSE.2017.06.031, 2017.

Graesser, J., Stanimirova, R., Tarrio, K., Copati, E. J., Volante, J. N., Verón, S. R., Banchero, S., Elena, H., Abelleyra, D. de, and Friedl, M. A.: Temporally-Consistent Annual Land Cover from Landsat Time Series in the Southern Cone of South America, Remote Sens (Basel), 14, https://doi.org/10.3390/rs14164005, 2022.

Gray, J., Sulla-Menashe, D., and Friedl, M. A.: User Guide to Collection 6.1 MODIS Land Cover Dynamics (MCD12Q2) Product, https://lpdaac.usgs.gov/documents/1417/MCD12Q2_User_Guide_V61.pdf (accessed 16 May 2023), 2022.

Guo, D., Lintern, A., Angus Webb, J., Ryu, D., Bende-Michl, U., Liu, S., and William Western, A.: A data-based predictive model for spatiotemporal variability in stream water quality, Hydrol Earth Syst Sci, 24, https://doi.org/10.5194/hess-24-827-2020, 2020.

Halefom, A., Sisay, E., Khare, D., Singh, L., and Worku, T.: Hydrological modeling of urban catchment using semi-distributed model, Model Earth Syst Environ, 3, 683–692, https://doi.org/10.1007/s40808-017-0327-7, 2017.

Hales, R. C., Williams, G. P., Nelson, E. J., Sowby, R. B., Ames, D. P., and Lozano, J. L. S.: Bias Correcting Discharge Simulations from the GEOGloWS Global Hydrologic Model, J Hydrol (Amst), 130279, https://doi.org/10.1016/j.jhydrol.2023.130279, 2023.

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., and Hobart, G. W.: Disturbance-Informed Annual Land Cover Classification Maps of Canada's Forested Ecosystems for a 29-Year Landsat Time Series, Canadian Journal of Remote Sensing, 44, https://doi.org/10.1080/07038992.2018.1437719, 2018.

Jin, S., Homer, C., Yang, L., Danielson, P., Dewitz, J., Li, C., Zhu, Z., Xian, G., and Howard, D.: Overall Methodology Design for the United States National Land Cover Database 2016 Products, Remote Sensing 2019, Vol. 11, Page 2971, 11, 2971, https://doi.org/10.3390/RS11242971, 2019.

Kalin, L., Isik, S., Schoonover, J. E., and Lockaby, B. G.: Predicting Water Quality in Unmonitored Watersheds Using Artificial Neural Networks, J Environ Qual, 39, https://doi.org/10.2134/jeq2009.0441, 2010.

Kaufman, Y. J.: Atmospheric Effects On Remote Sensing Of Surface Reflectance, https://doi.org/10.1117/12.966238, 0475, 20–33, https://doi.org/10.1117/12.966238, 1984.

Khodaee, M., Hwang, T., Ficklin, D. L., and Duncan, J. M.: With warming, spring streamflow peaks are more coupled with vegetation green-up than snowmelt in the northeastern United States, Hydrol Process, 36, e14621, https://doi.org/10.1002/HYP.14621, 2022.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., van den Broeke, M., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Val Martin, M., and Zeng, X.: The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty, J Adv Model Earth Syst, 11, https://doi.org/10.1029/2018MS001583, 2019.

Li, D., Bou-Zeid, E., Barlage, M., Chen, F., and Smith, J. A.: Development and evaluation of a mosaic approach in the WRF-Noah framework, Journal of Geophysical Research Atmospheres, 118, https://doi.org/10.1002/2013JD020657, 2013.

Liu, D. and Cai, S.: A Spatial-Temporal Modeling Approach to Reconstructing Land-Cover Change Trajectories from Multi-temporal Satellite Imagery, Annals of the Association of American Geographers, 102, https://doi.org/10.1080/00045608.2011.596357, 2012.

Li, Y., Chang, J., Luo, L., Wang, Y., Guo, A., Ma, F., and Fan, J.: Spatiotemporal impacts of land use land cover changes on hydrology from the mechanism perspective using SWAT model with time-varying parameters, Hydrology Research, 50, 244–261, https://doi.org/10.2166/NH.2018.006, 2019.

Luo, M., Li, F., Hao, D., Zhu, Q., Dashti, H., and Chen, M.: Uncertain spatial pattern of future land use and land cover change and its impacts on terrestrial carbon cycle over the Arctic–Boreal region of North America, Earths Future, 11, e2023EF003648, 2023.

Marufah, U., June, T., Faqih, A., Ali, A. A., Stiegler, C., and Knohl, A.: Implication of land use change to biogeophysical and biogeochemical processes in Jambi, Indonesia: Analysed using CLM5, Terrestrial, Atmospheric and Oceanic Sciences, 32, https://doi.org/10.3319/TAO.2020.12.17.01, 2021.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W. Van, Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, Trans ASABE, 50, 885–900, https://doi.org/10.13031/2013.23153, 2007.

Murphy, J. C.: Changing suspended sediment in United States rivers and streams: Linking sediment trends to changes in land use/cover, hydrology and climate, Hydrol Earth Syst Sci, 24, 991–1010, https://doi.org/10.5194/HESS-24-991-2020, 2020.

Myers, D. T., Ficklin, D. L., Robeson, S. M., Neupane, R. P., Botero-Acosta, A., and Avellaneda, P. M.: Choosing an arbitrary calibration period for hydrologic models: How much does it influence water balance simulations?, Hydrol Process, 35, e14045, https://doi.org/10.1002/hyp.14045, 2021a.

Myers, D. T., Ficklin, D. L., and Robeson, S. M.: Incorporating rain-on-snow into the SWAT model results in more accurate simulations of hydrologic extremes, J Hydrol (Amst), 603, 126972, https://doi.org/10.1016/J.JHYDROL.2021.126972, 2021b.
Nguyen, T. V, Dietrich, J., Dang, T. D., Tran, D. A., Van Doan, B., Sarrazin, F. J., Abbaspour, K., and Srinivasan, R.: An interactive graphical interface tool for parameter calibration, sensitivity analysis, uncertainty analysis, and visualization for the Soil and Water Assessment Tool, Environmental Modelling & Software, 156, 105497, https://doi.org/https://doi.org/10.1016/j.envsoft.2022.105497, 2022.
Qin, Y., Wang, D., Cao, Y., Cai, X., Liang, S., Beck, H. E., and Zeng, Z.: Sub-Grid Representation of Vegetation Cover in Land Surface Schemes Improves the Modeling of How Climate Responds to Deforestation, Geophys Res Lett, 50, https://doi.org/10.1029/2023GL104164, 2023.
Radeloff, V. C., Roy, D. P., Wulder, M. A., Anderson, M., Cook, B., Crawford, C. J., Friedl, M., Gao, F., Gorelick, N., and Hansen, M.: Need and vision for global medium-resolution Landsat and Sentinel-2 data products, Remote Sens Environ, 300, 113918, https://doi.org/10.1016/j.rse.2023.113918, 2024.
Rumora, L., Miler, M., and Medak, D.: Impact of Various Atmospheric Corrections on Sentinel-2 Land Cover Classification Accuracy Using Machine Learning Classifiers, ISPRS International Journal of Geo-Information 2020, Vol. 9, Page 277, 9, 277, https://doi.org/10.3390/IJGI9040277, 2020.
Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour Res, 46, https://doi.org/10.1029/2008WR007327, 2010.
Sexton, J. O., Song, X. P., Huang, C., Channan, S., Baker, M. E., and Townshend, J. R.: Urban growth of the Washington, D.C.-Baltimore, MD metropolitan region from 1984 to 2010 by annual, Landsat-based estimates of impervious cover, Remote Sens Environ, 129, https://doi.org/10.1016/j.rse.2012.10.025, 2013.
Sulla-Menashe, D. and Friedl, M. A.: User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product, https://doi.org/10.5067/MODIS/MCD12Q1, 2018.
Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., and Friedl, M. A.: Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product, Remote Sens Environ, 222, 183–194, 2019.
Vrugt, J. A. and Robinson, B. A.: Improved evolutionary optimization from genetically adaptive multimethod search, Proceedings of the National Academy of Sciences, https://doi.org/10.1073/pnas.0610471104, 2007.
Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J., and Rowe, C. M.: Statistics for the evaluation and comparison of models, J Geophys Res, 90, 8995, https://doi.org/10.1029/JC090iC05p08995, 1985.
Zhang, H. K., Roy, D. P., Yan, L., Li, Z., Huang, H., Vermote, E., Skakun, S., and Roger, J. C.: Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences, Remote Sens Environ, 215, 482–494, https://doi.org/10.1016/J.RSE.2018.04.031, 2018.

Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., Li, Y., Toman, E., Mallick, B., Zhang, X., and Brown, M.: Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm, Remote Sens Environ, 232, https://doi.org/10.1016/j.rse.2019.04.034, 2019.

Zhu, Z. and Woodcock, C. E.: Continuous change detection and classification of land cover using all available Landsat data, Remote Sens Environ, 144, https://doi.org/10.1016/j.rse.2014.01.011, 2014.

Zhu, Z., Woodcock, C. E., and Olofsson, P.: Continuous monitoring of forest disturbance using all available Landsat imagery, Remote Sens Environ, 122, https://doi.org/10.1016/j.rse.2011.10.030, 2012.