Reviewer 1:

Review – egusphere-2023-1171

Myers et al : Seasonal variation in landcover estimates reveals sensitivities and opportunities for environmental models

The authors present results from a study exploring the impact of using temporally resolved land cover data for water quality and hydrological model development. The manuscript reads reasonably well but there are some structural issues that need to be addressed (see specific comments below). My other concerns are relatively minor and are related to the presentation and discussion of results. I believe this manuscript could represent a useful contribution to our understanding of model uncertainty associated with seasonal landcover changes. While, the manuscript, in current form is not suitable for publication, with relatively minor revisions this study would be suitable and of broad interest to the readership of HESS.

> Response: We thank the reviewer for the helpful and constructive feedback. We have made improvements to address the structural and minor concerns below. Thank you for taking the time to help us communicate our findings, which we agree are of broad interest to HESS readership.

Specific comments:

Line 16: Qualify the potential implications on land cover characterization here.

> Response: We expanded on this abstract sentence to incorporate the implications of seasonal land cover characterization in models:
>
> “*Most readily available landuse/landcover (LULC) data are developed using growing season remote sensing images often at annual time steps, but LULC characteristics can have seasonal inconsistencies, which could impact geospatial models applied to another season of data.*” (to be updated at page 1, line 16 of the preprint)

Line 55: It seems like you need a research question/hypothesis linked to the spatiotemporal variability in land cover quantification. Where certain types of catchments more likely to exhibit large seasonal shifts in land cover quantification?

> Response: We agree that having a hypothesis here related to the variability would make communication of our structure more clear. We added the following text:

> "*We hypothesized that watersheds with mixed landcover types (e.g., a combination of built and trees) would have the greatest variability in landcover classification between growing and non-growing seasons due to heightened temporal inconsistencies, which could carry over into sensitivities for watershed-scale geospatial models.*" (to be added at page 2, line 57 of the preprint)

Figure 3. The differences in landcover between seasons seem most pronounced for 2016. This is interesting and isn't explored in the manuscript. This is important as you are using 2016 (i.e. national landcover database) as the reference and if this was a particularly anomalous year could there be implications for your conclusions? Perhaps think about exploring the drivers of inter annual variability here (e.g. climate drivers).

> <u>Response:</u> We thank the reviewer for pointing out this implication that was in need of further discussion to ensure the usability of our findings in different time periods. We now relate the 2016 LULC differences to those of other years, in particular the next instance of NLCD data available in 2019. We also now discuss potential causes of year-to-year changes based on previous work, such as uncorrected differences in atmospheric conditions or reflectance.
>
> "*The differences between seasons were not limited to a single year of data or watershed and could be more or less pronounced depending on the watershed and time period. For instance, our study watershed for the LULC change simulation (Case #2, Rock Creek) showed a 9% increase in built LULC, and a 12% decrease in tree area, in non-growing season relative to growing season Dynamic World data from 2016. Meanwhile, our study watershed for the independently calibrated models (Case #3, Difficult Run) showed a 12% decrease in tree cover and a 10% increase in built areas in the non-growing season compared to the growing season Dynamic World 2016. Over the entire time period of available Dynamic World estimates for these watersheds, growing season LULC estimates generally had more tree area, while non-growing season had more built area, and 2016 had the most pronounced differences (Figure 3c-f). For 2019, when the next instance of NLCD is available for comparisons, differences between non-growing and growing season estimates would be less pronounced for the Rock Creek Watershed (+5% built area and -8% trees), but approximately the same as 2016 for the Difficult Run watershed (+10% built area and -11% trees). In some years such as 2017-2018 the relationship could be reversed. 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.*" (to be updated at page 9, lines 176-184 of preprint)

Table 2. I'm not sure on the relevance of presenting the AIC score here? This is normally used for model selection –wouldn't the RMSE of the fitted values be a more useful indicator of differences between the land cover quantification methods.

> Response: We switched our AIC calculations to RMSE to better indicate model performance with different LULC seasons. Our conclusions did not change but are now better supported.
>
> "*The model for growing season built LULC vs. median SC had an R^2 of 0.69, while the same model for non-growing season LULC had an R^2 of 0.70, and the RMSE's for both models were within 3 RMSE units (150.16 and 148.08, respectively), which suggests similar performance. For perspective, a model created with developed classes from NLCD 2016 had a similar fit as both seasonal models (R^2 of 0.66 and RMSE of 155.91; Table 2), supporting that Dynamic World could be relevant for identifying LULC forcings affecting water quality particularly where regional products such as NLCD are not available.*" (to be updated at page 9, lines 194-199 of preprint)

Figure 4. Colour contrast makes it difficult to view the different lines/points. Perhaps consider using a different palette with stronger contrasts? Also this looks like a quadratic relationship rather than linear?

> Response: We modified the color palette of Figure 4 to be more contrasting, as well as changing the widths and types of lines. We also explored quadratic fits to the models instead of linear and the impacts to R2 and RMSE. The quadratic models had a better fit as the reviewer pointed out; thus, we updated the manuscript with the quadratic models.

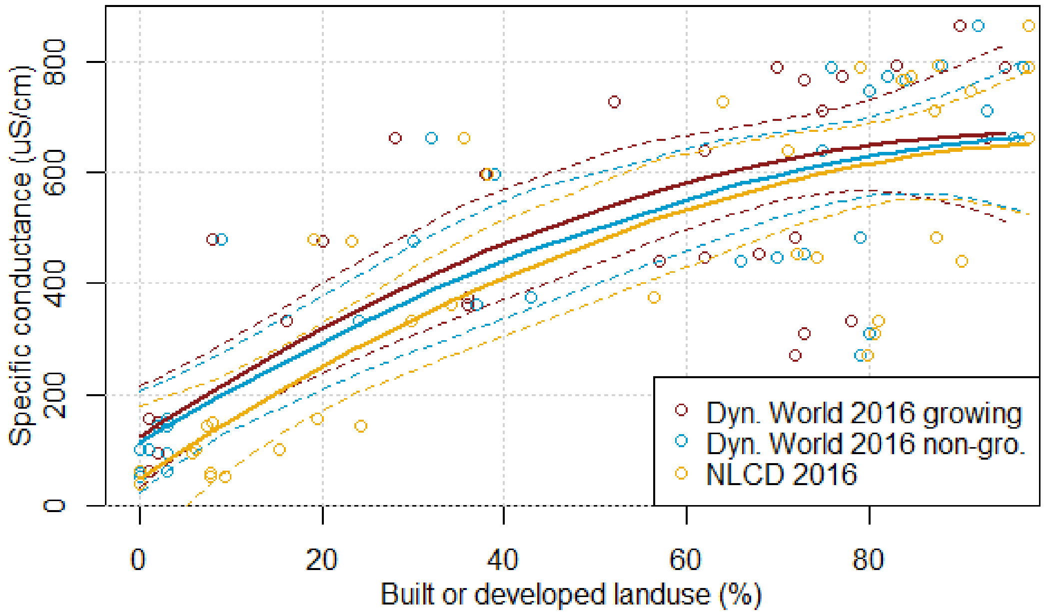

Figure 4: Modeled median specific conductance (SC) for 37 watersheds comparing Dynamic World 2016 growing and non-growing season built and NLCD 2016 developed LULC, with 95% confidence intervals as dashed lines.

Line 265: Move figure from supplementary material to support discussion of the implications for model parameters. I think this is an important part of the manuscript and should be given more prominence.

> Response: We moved the figure to the manuscript to support the discussion as suggested (now Figure 7).

Conclusion – avoid excessive referencing to other studies in the conclusion section.

> Response: We removed references from our conclusions section and instead ensured they were cited in the introduction and/or discussions. We believe the conclusion now reads easier following the reviewer's suggestion.
>
> "*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. The cause in temperate watersheds is primarily a sensitivity to changes from built to forest LULC proportions that affect modeled runoff and nutrient yields, representing temporal classification*

> *inconsistencies rather than actual succession or restoration. 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 Dynamic World LULC data in a model. The seasonal variation in Dynamic World LULC data we identified is pertinent for environmental models of future climates, biodiversity, habitat loss, land management, ecology, and biogeochemistry that are dependent on precise assessments of LULC change that could be affected by the seasonal classification variation.*" (to be updated at page 14, lines 277-288 of preprint)

Lines 300 -315: Rather than a bullet point list I suggest you develop a more coherent narrative focused on the implications of your findings and future research directions. This could go before the conclusion section.

> Response: We have now added a future research section (3.5) to the end of our discussion and replaced the bullets with a one-sentence summary in our conclusion.
>
> "*The impacts of seasonal landcover inconsistencies on geospatial models could yield several additional future research directions that build upon our findings. As our study used watershed-scale water quality and quantity investigations, further work should investigate how seasonal LULC classification inconsistencies could affect assessments of habitat, biodiversity, land management, ecology, global hydrology, and future climate based on LULC change (e.g., Yang et al., 2022; Di Vittorio et al., 2018; Hales et al., 2023). It may be particularly useful to explore whether the high resolution, high frequency LULC data could be used in LULC change models (e.g., Hood et al., 2021) to improve the temporal precision of interpolations between discrete LULC images. Future work could also investigate how seasonal LULC classification inconsistencies influence models outside our temperate study area (e.g., mountainous, arid, tropical, high-latitude, savannah, Mediterranean, continental) to gain a broader understanding of global geospatial model impacts. The use of high-frequency monitoring data (Zhang et al., 2023) could be explored to investigate the influence of high temporal resolution LULC on water quality patterns, as well as whether a modification to environmental models such as time varying parameters (Li et al., 2019) could account for the seasonal differences in Dynamic World LULC classifications. Future research could also incorporate LULC pixel probabilities from the Dynamic World dataset (Brown et al., 2022; Small and Sousa, 2023) into geospatial models and investigate their utility for environmental fields. Post-processing approaches for high temporal resolution LULC products to address seasonal inconsistencies (Sexton et al., 2013; Liu and Cai, 2012; Hermosilla et al., 2018) could aid in alleviating the impacts of seasonal inconsistencies causing model sensitivities as well. Finally, future work could investigate which seasons of LULC data are most accurate for different purposes, such as vegetation or impervious surface*

classification, and how causes of year-to-year inconsistencies in seasonal LULC estimates could affect models." (to be added after page 13, line 275 of preprint)

"We discussed future research directions which could advance capabilities to use high spatiotemporal resolution global LULC information such as Dynamic World for geospatial models across disciplines." (to be updated at page 14, line 295 to page 15, line 314 of preprint)

References:

Brown, C. F., Brumby, S. P., Guzder-Williams, B., et al. : 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.

Hales, R. C., Williams, G. P., Nelson, E. J., et al.: 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., et al.: 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.

Hood, R. R., Shenk, G. W., Dixon, R. L., et al.: The Chesapeake Bay program modeling system: Overview and recommendations for future development, Ecol Modell, 456, 109635, https://doi.org/10.1016/J.ECOLMODEL.2021.109635, 2021.

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.

Li, Y., Chang, J., Luo, L., Wang, Y., et al.: 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.

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.

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.

Sexton, J. O., Song, X. P., Huang, C., et al.: 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.

Small, C. and Sousa, D.: Spectral Characteristics of the Dynamic World Land Cover Classification, Remote Sens (Basel), 15, 575, 2023.

Yang, X., Rode, M., Jomaa, S., et al.: Functional Multi-Scale Integration of Agricultural Nitrogen-Budgets Into Catchment Water Quality Modeling, Geophys Res Lett, 49, https://doi.org/10.1029/2021GL096833, 2022.

Di Vittorio, A. V., Mao, J., Shi, X., et al.: Quantifying the Effects of Historical Land Cover Conversion Uncertainty on Global Carbon and Climate Estimates, Geophys Res Lett, 45, https://doi.org/10.1002/2017GL075124, 2018.

Zhang, H. K., Roy, D. P., Yan, L., et al.: 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.

Zhang, K., Bin Mamoon, W., Schwartz, E., and Parolari, A. J.: Reconstruction of Sparse Stream Flow and Concentration Time-Series Through Compressed Sensing, Geophys Res Lett, 50, e2022GL101177, https://doi.org/10.1029/2022GL101177, 2023.