Seasonal variation in landcover estimates reveals sensitivities and opportunities for environmental models

Daniel T. Myers[1*], David Jones[2], Diana Oviedo-Vargas[1], John Paul Schmit[2], Darren L. Ficklin[3], Xuesong Zhang[4]

[1] Stroud Water Research Center, 970 Spencer Road, Avondale, Pennsylvania 19311, USA
[2] National Park Service National Capital Region Network, 4598 MacArthur Blvd. NW, Washington, DC 20007, USA
[3] Department of Geography, Indiana University Bloomington, Student Building 120, 701 E. Kirkwood Avenue, Bloomington, IN 47405, USA
[4] Hydrology and Remote Sensing Laboratory, United States Department of Agriculture Agricultural Research Service, Bldg. 007, Rm. 104, BARC-West, Beltsville, MD 20705-2350, USA

Correspondence to: Daniel T. Myers (dmyers@stroudcenter.org)

Abstract. Most readily available landuse/landcover (LULC) data are developed using growing season remote sensing images often at annual time steps, ~~but seasonal changes in remote sensing data can lead to inconsistencies in LULC classification~~but LULC characteristics can have seasonal inconsistencies, which could impact geospatial models ~~applied to another season of data~~based on LULC. We used the Dynamic World near real-time global LULC dataset to compare how geospatial environmental models of water quality and hydrology respond to LULC estimated from growing vs. non-growing season ~~data~~LULC for temperate watersheds of the eastern United States. 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 ~~an increase~~differences in watershed nitrogen yields simulated by the Soil and Water Assessment Tool. Within reason, using separate calibration for each season may compensate for these inconsistencies, but lead to different model parameter optimizations. Our findings provide guidelines on the use of near real-time and high temporal resolution LULC in geospatial models.

1 Introduction

Environmental models incorporating landuse/landcover (LULC) data are common in many fields including hydrology, biogeochemistry, ecology, and climate science, often with decision-making implications (Hu et al., 2021; Baumgartner and Robinson, 2017; Naha et al., 2021; Li et al., 2021). 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). Characteristics of LULC (e.g., canopy density and precipitation interception) vary seasonally, particularly in temperate regions where vegetation leaf

cover is reduced during the non-growing season compared to the growing season (van Beusekom et al., 2014). This has prompted popular hydrological models such as the Soil and Water Assessment Tool (SWAT; Arnold et al., 1998) to include seasonal cycles for factors like leaf area and crops (Nkwasa et al., 2020; Frans et al., 2013). However, there can also be temporal inconsistencies in LULC classifications due to variation in spectral signals that are often not accounted for, such as built LULC being classified as other types within the course of a year, or other classes being classified as trees too quickly for natural succession (Cai et al., 2014; Gómez et al., 2016). 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). 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).

Present day high temporal resolution LULC datasets, such as the global Dynamic World (Brown et al., 2022), can facilitate the study of non-growing season and near real-time impacts of LULC classifications on environmental models, including those of hydrology and water quality. Dynamic World, which has a 10 m spatial resolution at 5-day intervals from Sentinel-2 satellites (2A and 2B), has comparable classification accuracy to other LULC datasets including the NLCD, European Space Agency World Cover, and ESRI Land Cover data (Venter et al., 2022; Brown et al., 2022), and its 5-day temporal resolution is much more frequent than the annual-or-longer frequency of other common LULC datasets. 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). Evaluation of LULC products at high spatiotemporal resolution is an important research need with vast societal implications (Radeloff et al., 2024).

Worldwide, investigations of LULC impacts to hydrology and water quality often employ regression-based models (Fu et al., 2019; Dow and Zampella, 2000), SWAT models simulating LULC change (Ni et al., 2021; Tong et al., 2009), and/or SWAT model configurations compared objectively to evaluate model performance (Fuka et al., 2012; Li et al., 2019). 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). We then used a long-term United States National Park Service (NPS) water quality dataset for temperate watersheds in the eastern United States, along with the above hydrologic and water quality models, to assess the use of seasonally-based LULC classificationsdata as an input for three modeling cases ranging from low to high complexity. We

asked "How different are model outputs (effect sizes) when using growing vs. non-growing season LULC inputs?" and "Are there differences in calibrated model performance if growing vs. non-growing season LULC input is used?" 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.

2 Materials and Methods

2.1 Study area and data

Our study area was 37 current (plus 18 historic) wadeable stream water quality sites monitored by the National Park Service National Capital Region Network (NCRN), with sites in Maryland, Virginia, West Virginia, and Washington DC, USA (Case #1; Figure 1). All sites are in the Chesapeake Bay watershed and were chosen to help inform natural resources management (Norris et al., 2011). This includes the 167 km^2 Rock Creek Watershed of Rock Creek National Park (Case #2) and the 150 km^2 Difficult Run Watershed of George Washington Memorial Parkway (Case #3), selected from the above watersheds for having continuous calibration and evaluation data.

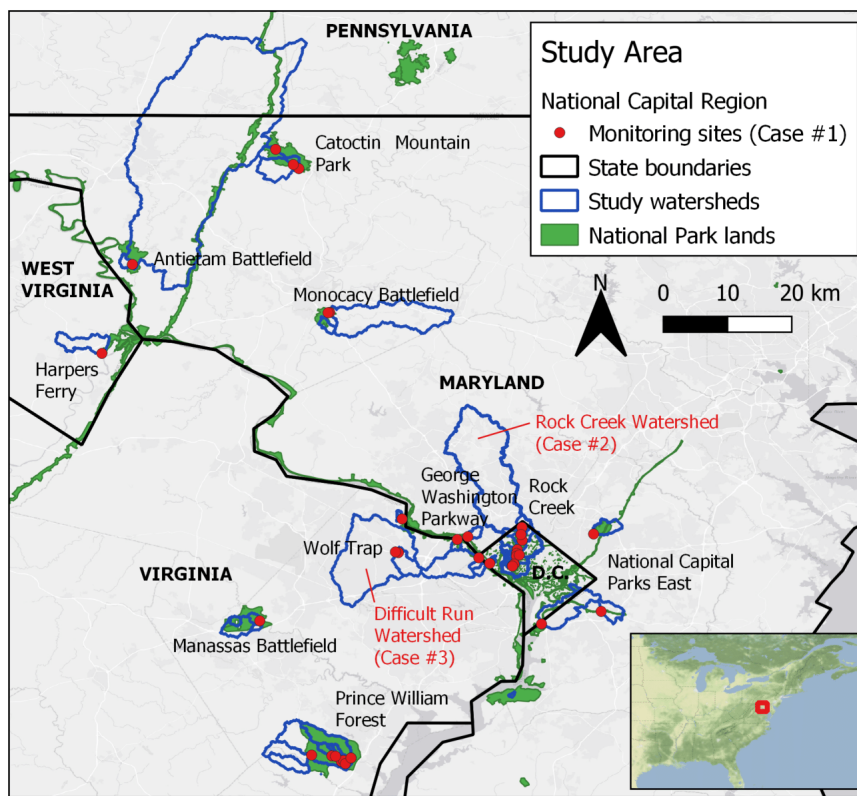



Figure 1: Study area map showing active monitoring sites and all (active + historic) watersheds.

Specific conductance (SC) can be used as an indicator of the overall amount of anthropogenic impacts to stream water quality in a watershed (Dow and Zampella, 2000). SC data from 2005-2018 for our study sites (Norris et al., 2011) were downloaded from the Water Quality Portal (https://www.waterqualitydata.us/; accessed 9 October, 2022). Discrete samples were taken every one to three months for each site following data quality controls and protocol (Norris et al., 2011), with an average of 179±89 measurements per site. Median SC values over the entire time period were used as the indicator of anthropogenic impacts to each stream for comparisons to compare water quality tendencies between monitoring sites (Dow and Zampella, 2000). Model calibration data are described in Sect. 2.5.

2.2 Seasonal landcover comparisons

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). 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, as developing a SWAT model requires the input of one LULC layer to input into the hydrologic and water quality models. Dynamic World's built class aggregates both hard structures (e.g., buildings and parking lots) and the surrounding vegetation, as is done in other common SWAT LULC inputs such as NLCD developed classes (Brown et al., 2022; Jin et al., 2019). We chose the years 2015-2016 because that was the earliest available Dynamic World data and nearest to the center of our 2005-2018 time period for water quality data, but repeated the process for every year of available Dynamic World data (2016-2021) for the Rock Creek and Difficult Run Watersheds to verify there was a seasonal cycle throughout years (see below). The timing of the data also aligned with the instance of NLCD data from 2016 for comparisons.

2.3 Experimental design

Different watersheds were tested in each case to demonstrate that the seasonal LULC estimate differences were not limited to a single watershed (Figure 2). For our water quality regressions (Case #1), we evaluated how well LULC classifications based on Dynamic World data from a single season could identify an LULC forcing affecting water quality at the watershed scale, following the common regression approach used in water quality investigations worldwide (Fu et al., 2019). We developed quadratic least-squares regression models of median stream SC values over the entire 2005-2018 period for 37 currently monitored NCRN sites explained by seasonal Dynamic World 2016 built LULC. The purpose for the water quality regressions case was to evaluate how well Dynamic World data could identify an LULC forcing affecting water quality at the watershed scale, following the common regression approach used in water quality investigations worldwide (Fu et al., 2019). Performance measures including R^2 and root mean square error (RMSE; Willmott et al., 1985) were used to compare models from different seasons. 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. Here, we developed and calibrated SWAT hydrologic and nitrogen (nitrate-N + nitrite-N) yield models for the Rock Creek Watershed, then used them to simulate an ~~LULC change~~ change in LULC classification between growing and non-growing seasons. ~~The purpose for the LULC change simulation case was to evaluate 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.~~ 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. Here, we developed and calibrated SWAT hydrologic models with growing and non-growing season Dynamic World 2016 inputs independently of one another for the Difficult Run Watershed. ~~The purpose for the independently calibrated models case was to assess the performance of seasonally tuned models rather than the single model of the land cover change case, to provide fairer comparison of calibrated model performances since each model was optimized to its unique LULC situation.~~ For each case we repeated the analysis with LULC from the commonly-used NLCD 2016 for comparison.

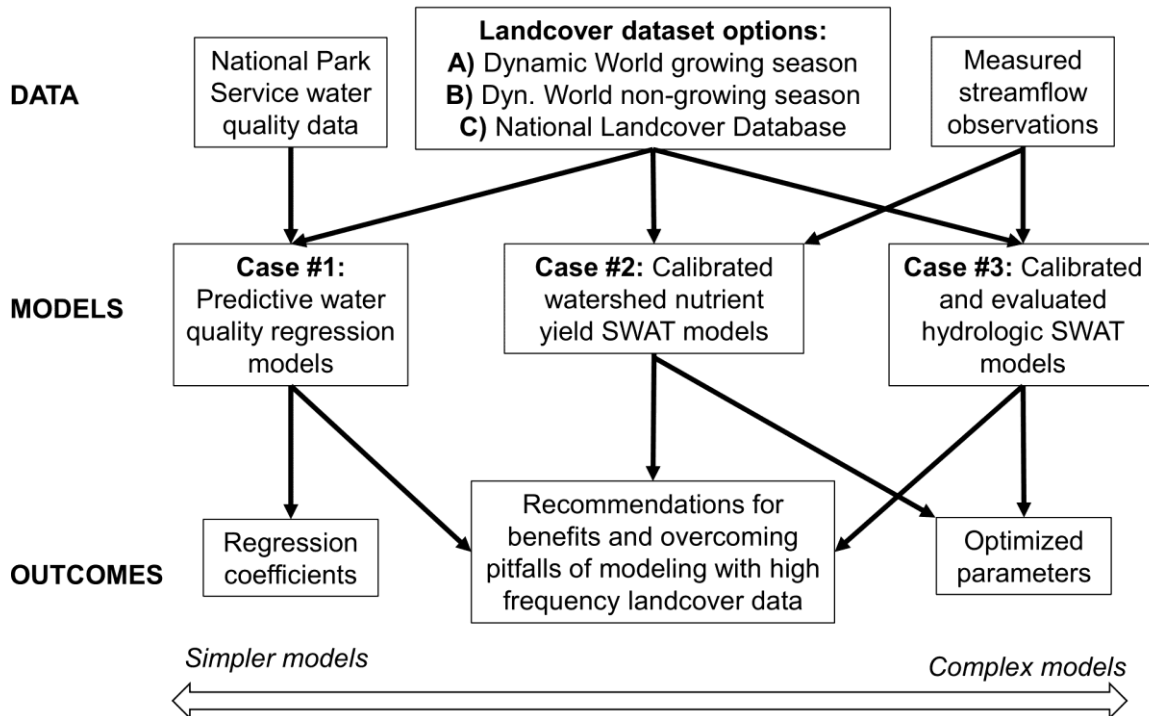



Figure 2: Conceptual diagram of the study.

2.4 Soil and Water Assessment Tool

135   SWAT is the most common water quality model globally (Fu et al., 2019) and has been used in over 6,000 peer-reviewed studies (https://www.card.iastate.edu/swat_articles/, accessed 7 January, 2024). The SWAT models (rev. 681) used in this study simulated streamflow using a water balance approach (Arnold et al., 1998, 2013), surface runoff using the runoff curve number (NRCS, 1986), groundwater flow using a water balance for shallow aquifer storage (Arnold et al., 1998), snowmelt based on snowpack temperature (Fontaine et al., 2002), and evapotranspiration using the Penman-Monteith method

140   (Monteith, 1965; Ritchie, 1972). Nitrogen yields were simulated based on estimates of runoff, crop use, lateral flow, percolation, and concentrations in soil and water (Arnold et al., 1998). 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

145   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 (Table S1) was input to generate tables that represent base watershed conditions (Abbaspour et al., 2019; Leeper et al., 2015; Lehner et al., 2006; Lindsay, 2022; Sugarbaker et al., 2014; USGS, 2022; USDA, 2022; Ries et al., 2017). We created a new SWAT LULC look-up table for QSWAT to read Dynamic World data and recreate HRUs (Table S2). The Rock Creek models for LULC change simulation (Case #2) had 13 subbasins, each

150   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

155   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. We chose the SWAT model for this study because it can be used to support water resource decision making in mixed-LULC watersheds (Koltsida et al., 2023).

2.5 Sensitivity analysis and calibration

160   The Rock Creek models (Case #2) used parameters calibrated with a Latin hypercube approach (to generate a large number of parameter sets; Abbaspour et al., 2004) to the SWAT model with growing season Dynamic World 2016 inputs, using R-SWAT software (Nguyen et al., 2022). R-SWAT is an open source, graphic interface, parallelizable, and user-friendly tool to calibrate the SWAT model and analyze results (Nguyen et al., 2022). 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

165   S3. 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. The years 2013-2021 were used in the simulations as these were the years the USGS station had been active for streamflow, and there was a 3 year model warm-up period (2010-2012) to reduce the influence of initial states. The calibrated parameter set was chosen as having the best performing Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970) values for streamflow and nitrogen yield out of the sample of parameter sets.

For Case #3, sensitivities of Difficult Run Watershed SWAT model performance to specific parameters were analyzed using the density-based PAWN method in the Sensitivity Analysis for Everybody (SAFE) toolbox (Pianosi and Wagener, 2015; Pianosi et al., 2015; Zadeh et al., 2017). Eight thousand SWAT model runs with growing season Dynamic World 2016 data were used for the sensitivity analysis, based on Myers et al. (2021a). We analyzed the sensitivity of 35 parameters and then chose the top 10 parameters with sensitivities greater than the dummy parameter to use in the calibration (Table 1). 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) and NSE as the objective function (the metric that the algorithm aims to maximize) and observed daily streamflow from USGS station 01646000 (with the first half for calibration and latter half for validation; Figure S1). In addition to NSE, metrics for Kling-Gupta Efficiency (KGE; Gupta et al., 2009) and refined Index of Agreement (d_r; Willmott et al., 2012) were calculated to confirm our interpretations, with higher values implying better model performance.

Table 1: Parameters used in SWAT model streamflow calibration for Difficult Run Watershed (Case #3), for models input with growing and non-growing season Dynamic World 2016 data, as well as the model with NLCD 2016 input. Further descriptions of these parameters can be found in Table S4.

Symbol	Definition †	Lower Limit	Upper Limit	Calibrated Growing	Calibrated Non-growing	Calibrated NLCD 2016
CH_KII.rte	Channel hydraulic conductivity (mm/h) (v)	0.1	150	0.11	3.86	0.14
ALPHA_BNK.rte	Bank flow recession constant (v)	0.01	1	0.14	0.27	1.00
CN_F.mgt	Runoff curve number (r)	-0.2	0.2	-0.17	-0.20	-0.08
SNO50COV.bsn	Fraction of SNOCOVMX for 50% cover (v)	0.01	0.8	0.03	0.03	0.25
ESCO.hru	Soil evaporation compensation coef. (v)	0.01	1	0.01	0.03	0.35
CH_NII.rte	Manning's n value for main channel (v)	0.01	0.30	0.30	0.30	0.30
SOL_BD.sol	Soil moist bulk density (r)	-0.2	0.2	-0.19	-0.01	0.00

SNOCOVMX.bsn	Snow depth above which is 100% cover (mm) (v)	0	500	471	496	205
SFTMP.bsn	Snowfall temperature threshold (°C) (v)	0	3	0.95	0.98	1.02
SOL_AWC.sol	Available Water Capacity (r)	-0.25	0.25	-0.23	-0.25	-0.23

† A 'v' indicates that the original parameter from QSWAT was replaced by the calibrated value globally, in the same unit. An 'r' indicates that the original parameter was modified relatively, multiplying it regionally by 1 + the calibrated value (e.g. a value of -0.2 reduces the original parameter by 20%).

3 Results and discussion

3.1 Seasonal landcover comparisons

The Dynamic World 2016 data classified a greater area of the 55 watersheds as trees during the growing season than during the non-growing season, typically by 5-10% of watershed area (Figure 3a). During the non-growing season, some areas classified as trees during the growing season were instead given built or shrubland LULC classes. Differences in seasonal LULC classifications in Dynamic World data were strongest in mixed-LULC watersheds (i.e., watersheds with 15% to 85% of the area classified as built LULC), and weaker in very low built or very high built percentage watersheds (R^2=0.49, df=52, F=24.82, p<0.001; Figure 3b). There was a relative mean absolute difference (RMAD) of 9.0% of watershed area between NLCD 2016 developed (including open space, low, medium, and high intensity) and Dynamic World 2016 growing season built data (5.9% using non-growing season built data) for the 37 currently monitored watersheds (Figure S2 and Table S5).

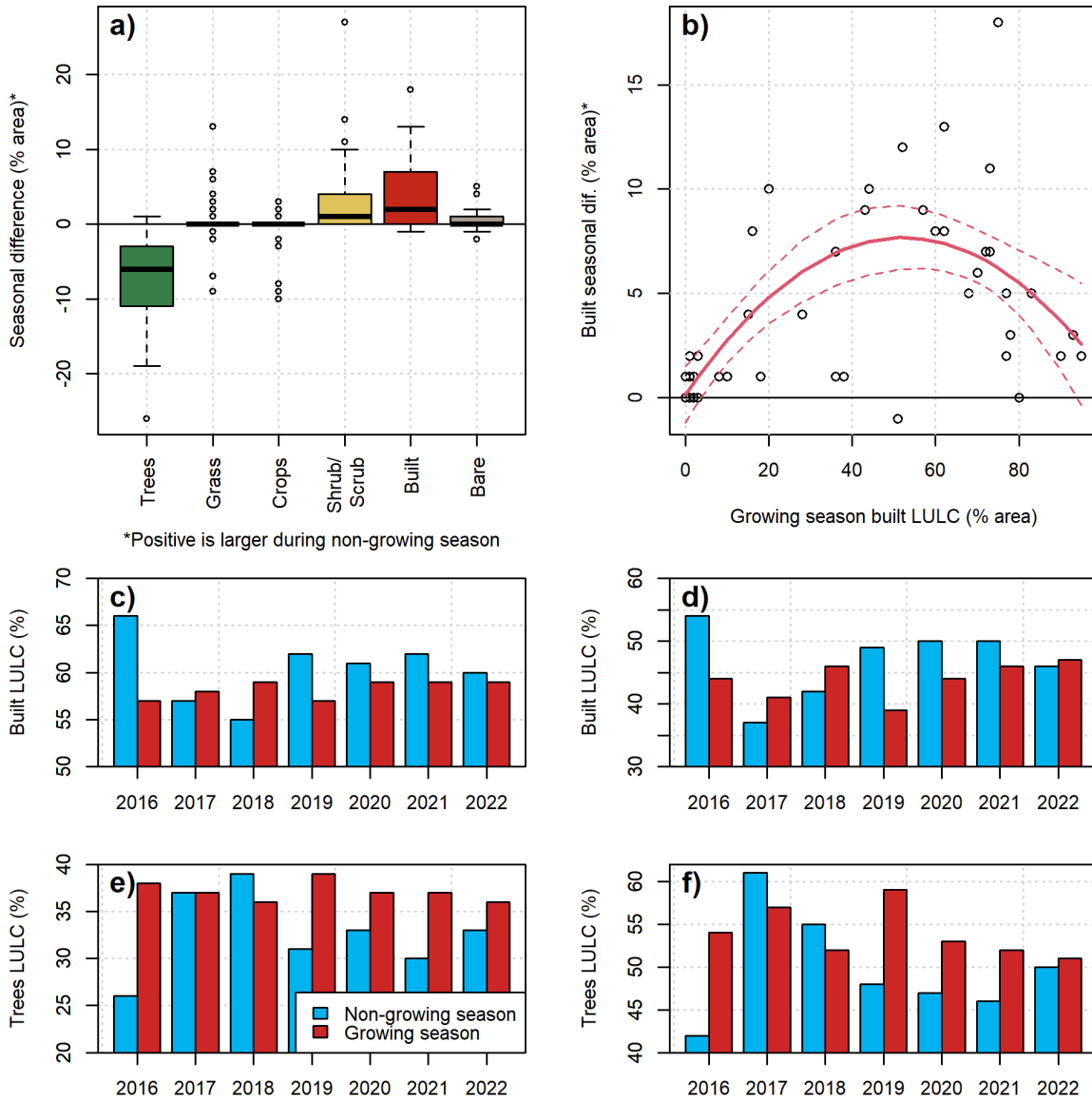



Figure 3: All using Dynamic World 2016: a) Difference between growing and non-growing season LULC for 55 watersheds (classes of water, flooded vegetation, barren, and snow/ice were approximately 1% of watershed area so omitted; boxplots show median, interquartile range (IQR), and outliers outside 1.5 * IQR), b) Quadratic relationship between built area and the seasonal difference in built area for 55 watersheds, with 95% confidence intervals as dashed lines, c) and d) Time series of built area estimates for the Rock Creek and Difficult Run Watersheds, respectively, and e,f) same as above but for tree area.

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.

Changes in LULC estimates between seasons were often concentrated along forested edges of mixed-LULC areas (Figure S3). In these deciduous areas, such as the edges of mixed residential/forested zones, leaf cover decreases during the non-growing season, which could be exposing other types of LULC underneath, or making forest more difficult to distinguish from surrounding built area for the classifications. 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).

3.2 Case #1: Water quality regressions

Median stream water specific conductance (SC) was positively correlated with 2016 Dynamic World built LULC during both seasons (Figure 4; Table 2). This relationship is expected and confirms that urban development has a strong positive effect on surface water salinization (Utz et al., 2022; Kaushal et al., 2005). 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.

Table 2: Regression models for specific conductance for the growing vs. non-growing seasons of Dynamic World 2016 built data and the NLCD 2016 developed classes model (df=34). CI: upper and lower 95% confidence intervals. Quadratic equation: $ax^2 + bx + c$.

LULC	a	b	c	R^2	F	p-value	CI (a)	CI (b)	RMSE
Dyn. World growing season	-0.05	10.83	123.65	0.69	37.52	<0.001	-0.13-0.02	4.58-17.07	150.16
Dyn. World non-growing season	-0.04	9.96	113.59	0.70	39.07	<0.001	-0.11-0.02	3.70-16.21	148.08
NLCD 2016	-0.05	11.03	49.04	0.66	33.57	<0.001	-0.13-0.03	3.30-18.76	155.91

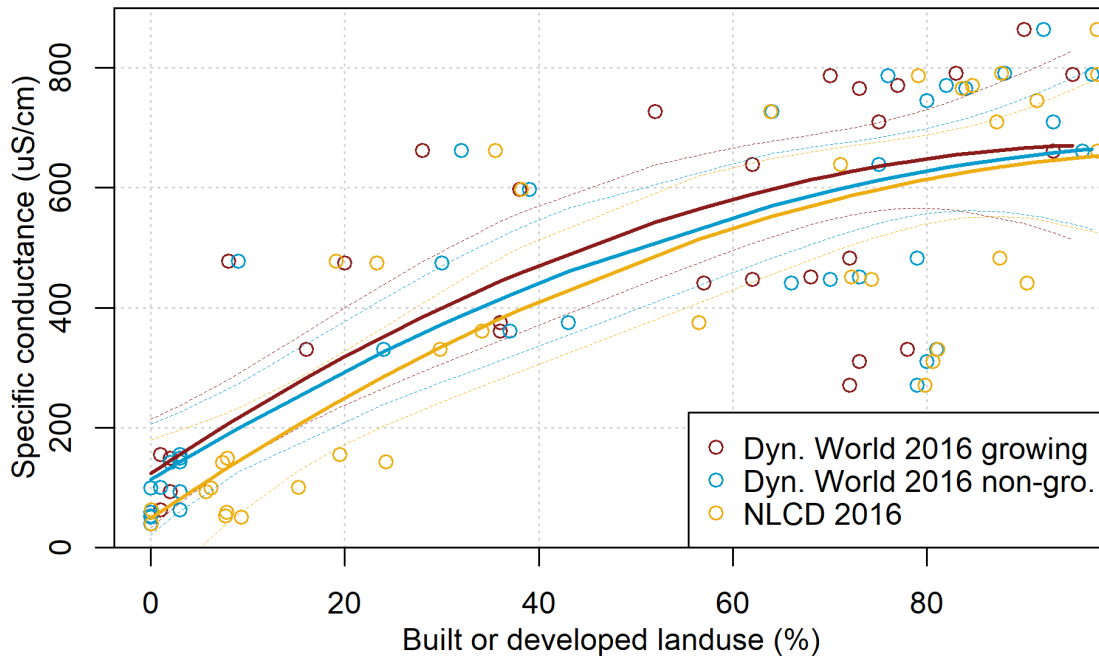



Figure 4: Modeled median specific conductance (SC) for 37 watersheds comparing Dynamic World 2016 growing and non-growing season built and NLCD 2016 developed LULC, with 95% confidence intervals as dashed lines.

3.3 Case #2: Hydrologic and nitrogen yield models

Our Rock Creek Watershed SWAT model for streamflow and nitrogen yield, developed and calibrated using Dynamic World 2016 growing season data, performed with a streamflow calibration NSE of 0.56 (validation NSE of 0.65), nitrogen yield calibration NSE of 0.45 (validation NSE of 0.80), and nitrogen yield calibration percent bias (PBIAS, where <0 implies overestimation bias; Gupta et al., 1999) of 14.6% (validation PBIAS of 1.6%) (Table 3). 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)Therefore, we concluded that the model developed with Dynamic World 2016 growing season data was reliably simulating real conditions at the monthly time step

(Figure 5a,b; red circles). 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 (Table 3; Figure 5a,b; blue circles). Note that both models were run over the same time period to compare performance. Also, the model simulated 50% greater nitrogen yield over the entire 2013-2021 time period when non-growing season Dynamic World 2016 data was used as the LULC input, rather than growing season LULC (Figure 5c). These discrepancies between model outputs are not negligible. In relative terms, this difference is greater than the current pollutant load reduction target for Chesapeake Bay of 17% total nitrogen load (Maryland Department of Environment, 2019). Therefore, we advise to take the potential seasonal variability of Dynamic World LULC estimates into consideration if used to design water quality improvement efforts, particularly when decision making is involved, or an LULC change is being simulated. A model could be fit to one season of LULC, but have bias if transferred to a different seasontime period of LULC estimates due to temporal inconsistencies. 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.

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).

SWAT LULC input	Period	Streamflow NSE	N yield NSE	N yield MAE (kg)	N yield PBIAS
Dyn. World 2016 growing season	Calibration	0.65	0.45	713	14.6%
Dyn. World 2016 growing season	Validation	0.56	0.80	909	1.6%
Dyn. World 2016 non-growing season	Calibration	0.35	-0.53	1177	-34.4%
Dyn. World 2016 non-growing season	Validation	0.21	-2.00	3205	-57.4%
NLCD 2016	Calibration	0.71	-1.14	1694	-7.8%
NLCD 2016	Validation	0.85	-0.33	2364	22.1%

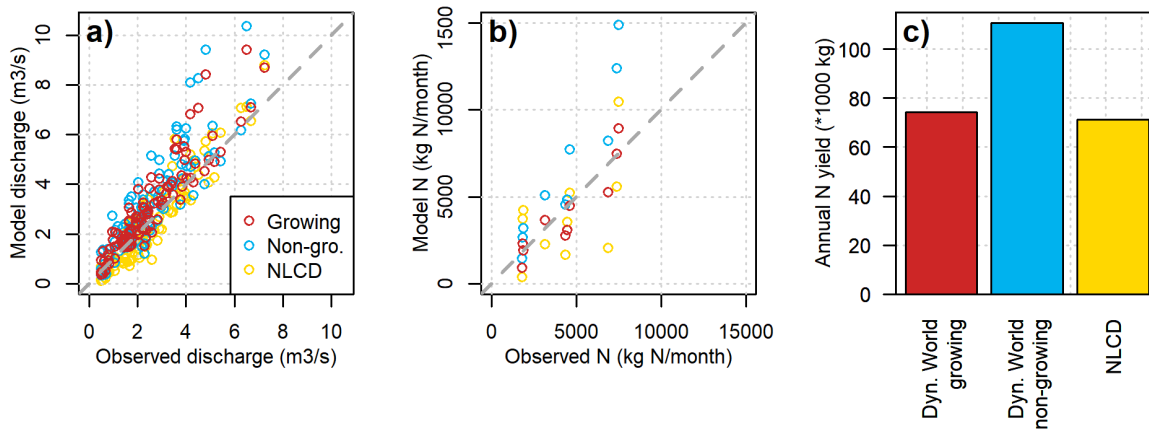



Figure 5: a) Observed vs. simulated monthly discharge for the Rock Creek Watershed comparing Dynamic World 2016 growing and non-growing season built and NLCD 2016 developed LULC, b) Same for monthly nitrogen (N) yields for Rock Creek, and c) Modeled average annual nitrogen yields for Rock Creek.

285

      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

290   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

295   NLCD 2016 LULC input using the same parameter adjustments (Figure 5c).

3.4 Case #3: Independently calibrated hydrologic models

      The individually calibrated SWAT models using growing season vs. non-growing season Dynamic World 2016 LULC input for the Difficult Run Watershed had comparable performance when simulating streamflow, despite the differences in LULC inputs (10% increase in built areas and 12% decrease in tree cover for the non-growing season LULC input). NSE

300   performance metrics at the daily time step were between 0.52 and 0.54 for each model with Dynamic World LULC over the calibration and validation time periods, KGE was between 0.61 and 0.75, and d_r (which by not squaring errors provides a better measure of low flow performance) only ranged between 0.68 and 0.70 (Table 4; scatterplots in log scale to show daily baseflows and time series are presented in Figure 6a-d). 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). For perspective, the

305 SWAT model calibrated with NLCD 2016 LULC had an NSE of 0.48 for the calibration period and 0.47 over the validation period (Table 4). 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). At the

310 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).

Table 4: Comparison of streamflow performance for calibrated SWAT models developed independently with Dynamic World 2016 growing
315 season LULC input, Dynamic World 2016 non-growing season LULC input, and NLCD 2016, at the daily time step for the Difficult Run Watershed (Case #3). Performance indices are R^2, NSE, Kling-Gupta Efficiency (KGE), and refined Index of Agreement (d_r).

SWAT landuse input	Period	R^2	NSE	KGE	d_r
Growing season	Calibration	0.54	0.53	0.61	0.69
Non-growing season	Calibration	0.54	0.54	0.65	0.70
NLCD 2016	Calibration	0.49	0.48	0.56	0.69
Growing season	Validation	0.56	0.53	0.73	0.68
Non-growing season	Validation	0.57	0.52	0.75	0.68
NLCD 2016	Validation	0.53	0.47	0.69	0.68

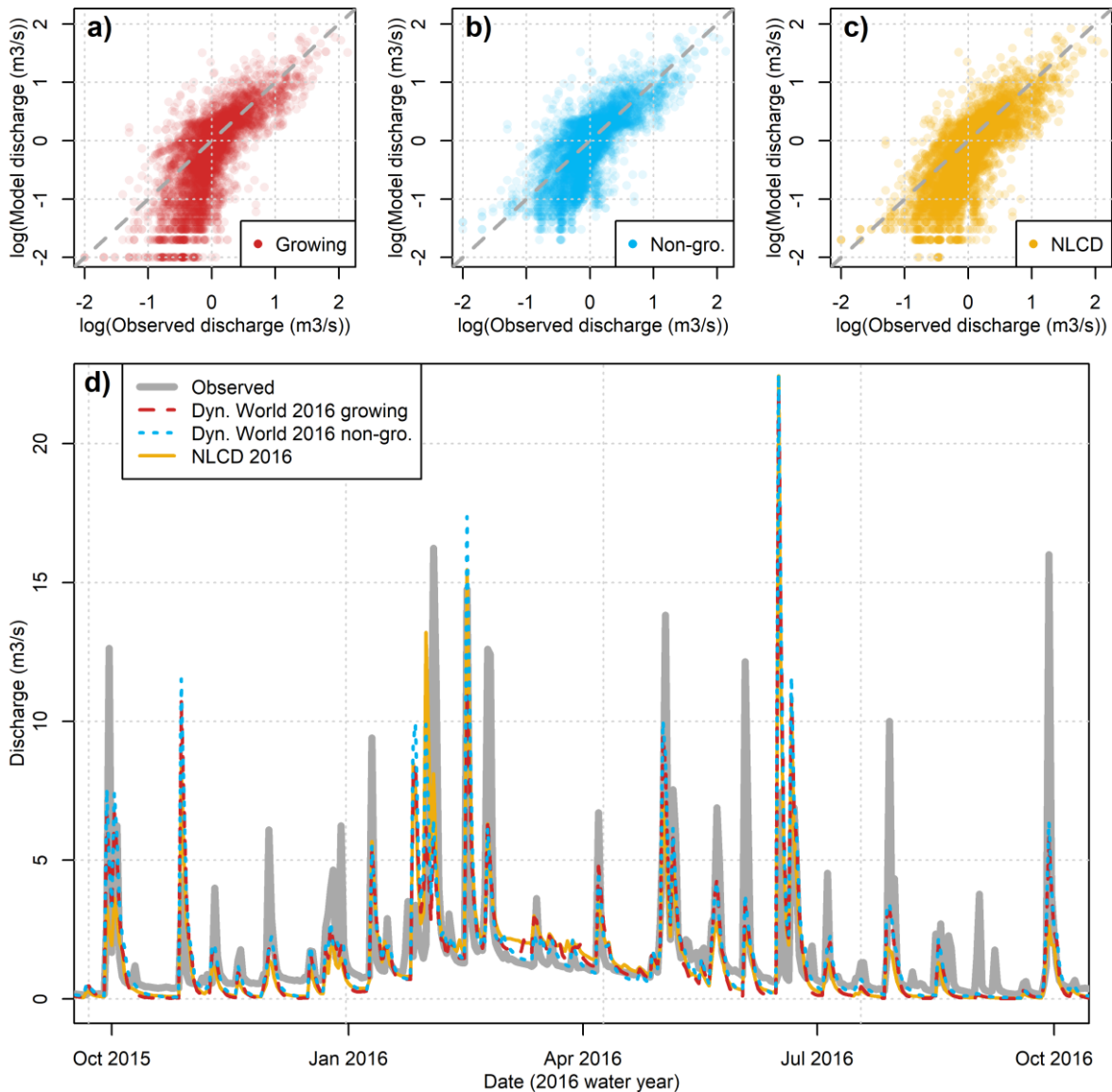



Figure 6: Daily discharge models for the Difficult Run Watershed displaying base-10 log, so that daily baseflows and low flows are visible, comparing independently calibrated models with a) Dynamic World 2016 growing season LULC, b) Dynamic World 2016 non-growing season LULC, and c) NLCD 2016. Also d) Time series of Difficult Run modeled discharge.

The most sensitive parameters for the Difficult Run Watershed case were channel hydraulic conductivity (CH_KII), bank flow recession coefficient (ALPHA_BNK), and runoff curve number (CN_F) (Figure 7). Among these and other sensitive parameters, there were differences in optimized values depending upon the SWAT LULC input (Table 1). For example, the CN_F adjustment optimized to -0.17 for growing season Dynamic World 2016, -0.20 for non-growing season Dynamic World

2016, and -0.08 for NLCD 2016 inputs, suggesting that the optimization adjusted runoff processes to compensate for the different proportions of LULC. The difference in forests of 12% of watershed area between growing and non-growing season Dynamic World 2016 data for Difficult Run (Table S5) is as large a difference as real changes in forests that have been found

330   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). 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). 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

335   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).

340

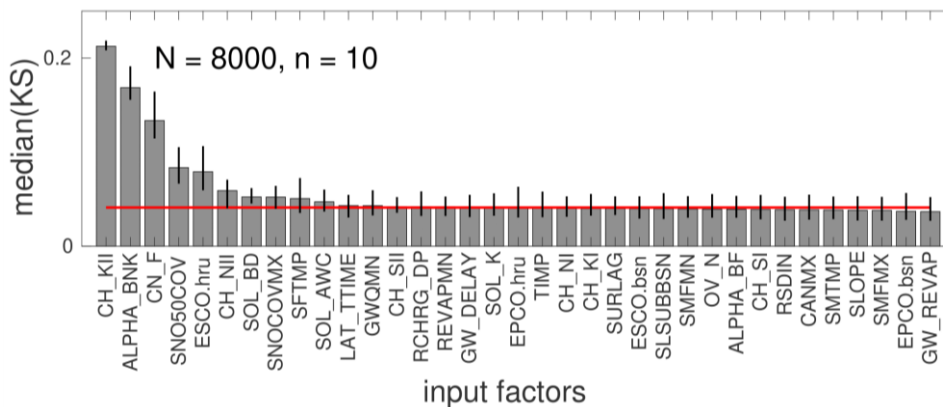



Figure 7. PAWN sensitivity analysis results ranking the SWAT parameters from most to least sensitive, using 8,000 samples (N) and conditioning intervals (n) of 10. The red line is the "dummy" parameter and bars are 95% confidence intervals. KS: Kolmogorov-Smirnov statistic. Higher median KS indicates higher sensitivity of SWAT model streamflow output to the parameter.

345   **3.5 Future directions**

Illogical LULC changes betweenclassifications related to seasonal differences in remote sensing data from different seasons 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 be a considerationconsidered during an LULC

350   update inwere an updated LULC layer to be used for a modelling approaches such as Hales et al. (2023), whereich bias corrected 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). 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. The strength of the effect of the illogical seasonal LULC change on the model outputs and optimized parameters would depend on many factors including model processes and spatiotemporal extent. A model intercomparison study in this regard would likely be a meaningful contribution to the advancement of the field into higher spatiotemporal capabilities.

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.

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 (Yang et al., 2022; Di Vittorio et al., 2018; Hales et al., 2023; Hood et al., 2021). 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,

390 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; Zhao et al., 2019) could aid

395 in alleviating the impacts of seasonal inconsistencies ~~causing~~ that cause 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.

4 Conclusions

When seasonal changes in LULC data occur, due to classification difficulties such as vegetation cycles (e.g.,

400 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 (Figure 8), 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

405 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. With a limited geographic scope (e.g., temperate watersheds) and

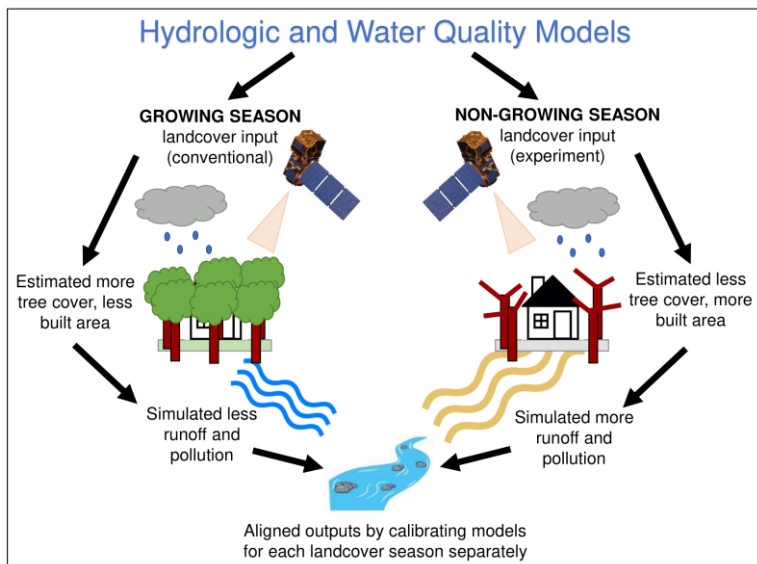
410 small sample of models, our work does not intend to show definitively when, where, or in what model configurations these sensitivities would occur, but that they are a possibility that modelers should be aware of. We discussed future research directions which could advance capabilities to use high spatiotemporal resolution global LULC information such as Dynamic World for geospatial models across disciplines.

415

Figure 8: Conceptual diagram of the conclusions of the study in temperate watersheds of the eastern United States.

Code and data availability

Data from this study, including the LULC images, water quality data, and model outputs from each case, are available from Mendeley Data at https://doi.org/10.17632/bbb9xbpv22.3 (Myers et al., 2022). Codes from this study, including Google

420 Earth Engine scripts and those to reproduce figures and analyses, are available on GitHub at https://github.com/Danmyers901/Calibration/tree/master/Landcover.

Supplementary information

Supplementary material for this article is available online for Figures S1-S53 and Tables S1-S5.

Author contribution

425 D.T.M. contributed to conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, and visualization. D.J. contributed to conceptualization, methodology, resources, writing – original draft, writing – review & editing, formal analysis, and investigation. D.O.V contributed to conceptualization, methodology, resources, writing – original draft, writing – review & editing, supervision, project administration, and funding acquisition. J.P.S. contributed to conceptualization, methodology, resources, writing –

430 original draft, writing – review & editing, formal analysis, and investigation. D.L.F. contributed to methodology, validation,

investigation, writing – original draft, and writing – review & editing. X.Z. contributed to methodology, validation, investigation, writing – original draft, and writing – review & editing.

Competing interests

The authors declare that they have no conflict of interest.

435 **Acknowledgments**

This work was supported by the National Park Service National Capital Region Network and Stroud Water Research Center. We also thank the National Science Foundation [grant number CNS-0521433], Indiana University Pervasive Technology Institute, Lilly Endowment, Inc., Indiana METACyt Initiative, and Shared University Research Grants from IBM, Inc. to Indiana University for programming and computing assistance. We thank Liz Matthews, Charles Wainright, Andrejs
440 Brolis, and Lindsay Ashley of the National Park Service for monitoring data and guidance. We also thank Jacob Price of the Stroud Water Research Center for help improving the communication of our findings, and Erin Hestir of the University of California Merced for help interpreting our results, and Wouter Knoben and the other peer reviewers for improving the clarity of our approach and novelties. We also give credit to Dr. Knoben for the idea we discuss to separate stable from changing HRUs. X.Z. is supported by U.S. Department of Agriculture - Agricultural Research Service.

445 **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,
450   2004.

Abbaspour, K. C., Vaghefi, S. A., Yang, H., and Srinivasan, R.: Global soil, landuse, evapotranspiration, historical and future weather databases for SWAT Applications, Scientific Data 2019 6:1, 6, 1–11, https://doi.org/10.1038/s41597-019-0282-4, 2019.

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

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large Area Hydrologic Modeling and Assessment Part I : Model Development, J Am Water Resour Assoc, 34, 73–89, https://doi.org/10.1111/j.1752-1688.1998.tb05961.x, 1998.

Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R., Haney, E. B., and Neitsch, S. L.: Soil & Water Assessment Tool: Input/output documentation. version 2012, Texas Water Resources Institute, TR-439, 2013.

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.

Baumgartner, S. D. and Robinson, C. T.: Changes in macroinvertebrate trophic structure along a land-use gradient within a lowland stream network, Aquat Sci, 79, 418, https://doi.org/10.1007/s00027-016-0506-z, 2017.

van Beusekom, A. E., Hay, L. E., Viger, R. J., Gould, W. A., Collazo, J. A., and Henareh Khalyani, A.: The Effects of Changing Land Cover on Streamflow Simulation in Puerto Rico, J Am Water Resour Assoc, 50, 1575–1593, https://doi.org/10.1111/jawr.12227, 2014.

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, https://doi.org/10.1016/J.EJRH.2022.101202, 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.

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.

Fontaine, T. A., Cruickshank, T. S., Arnold, J. G., and Hotchkiss, R. H.: Development of a snowfall-snowmelt routine for mountainous terrain for the soil water assessment tool (SWAT), J Hydrol (Amst), 262, 209–223, https://doi.org/10.1016/S0022-1694(02)00029-X, 2002.

Frans, C., Istanbulluoglu, E., Mishra, V., Munoz-Arriola, F., and Lettenmaier, D. P.: Are climatic or land cover changes the dominant cause of runoff trends in the Upper Mississippi River Basin?, Geophys Res Lett, 40, https://doi.org/10.1002/grl.50262, 2013.

495 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,

500 48, 1151–1161, https://doi.org/10.1111/j.1752-1688.2012.00680.x, 2012.

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.

Gómez, C., White, J. C., and Wulder, M. A.: Optical remotely sensed time series data for land cover classification: A review,

505 https://doi.org/10.1016/j.isprsjprs.2016.03.008, 2016.

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,

510 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

515 model for spatiotemporal variability in stream water quality, Hydrol Earth Syst Sci, 24, https://doi.org/10.5194/hess-24-827-2020, 2020.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration, J Hydrol Eng, 4, 135–143, https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135), 1999.

520 Gupta, H. v, Kling, H., Yilmaz, K. K., Martinez, G. F., and Kling, H.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J Hydrol (Amst), 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

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.

525   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
530   Remote Sensing, 44, https://doi.org/10.1080/07038992.2018.1437719, 2018.

Hood, R. R., Shenk, G. W., Dixon, R. L., Smith, S. M. C., Ball, W. P., Bash, J. O., Batiuk, R., Boomer, K., Brady, D. C., Cerco, C., Claggett, P., de Mutsert, K., Easton, Z. M., Elmore, A. J., Friedrichs, M. A. M., Harris, L. A., Ihde, T. F., Lacher, L., Li, L., Linker, L. C., Miller, A., Moriarty, J., Noe, G. B., Onyullo, G., Rose, K., Skalak, K., Tian, R., Veith, T. L., Wainger, L., Weller, D., and Zhang, Y. J.: The Chesapeake Bay program modeling system: Overview
535   and recommendations for future development, Ecol Modell, 456, 109635, https://doi.org/10.1016/J.ECOLMODEL.2021.109635, 2021.

Hu, X., Huang, B., Verones, F., Cavalett, O., and Cherubini, F.: Overview of recent land-cover changes in biodiversity hotspots, Front Ecol Environ, 19, https://doi.org/10.1002/fee.2276, 2021.

Jin, S., Homer, C., Yang, L., Danielson, P., Dewitz, J., Li, C., Zhu, Z., Xian, G., and Howard, D.: Overall Methodology Design
540   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,
545   20–33, https://doi.org/10.1117/12.966238, 1984.

Kaushal, S. S., Groffman, P. M., Likens, G. E., Belt, K. T., Stack, W. P., Kelly, V. R., Band, L. E., and Fisher, G. T.: Increased salinization of fresh water in the northeastern United States, Proceedings of the National Academy of Sciences, 102, 13517–13520, https://doi.org/10.1073/PNAS.0506414102, 2005.

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

Koltsida, E., Mamassis, N., and Kallioras, A.: Hydrological modeling using the Soil and Water Assessment Tool in urban and peri-urban environments: the case of Kifisos experimental subbasin (Athens, Greece), Hydrol Earth Syst Sci, 27, https://doi.org/10.5194/hess-27-917-2023, 2023.

555   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.

Leeper, R. D., Rennie, J., and Palecki, M. A.: Observational Perspectives from U.S. Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) Network: Temperature and Precipitation Comparison, J Atmos Ocean Technol, 32, 703–721, https://doi.org/10.1175/JTECH-D-14-00172.1, 2015.

Lehner, B., Verdin, K., and Jarvis, A.: HydroSHEDS Technical Documentation, https://data.hydrosheds.org/file/technical-documentation/HydroSHEDS_TechDoc_v1_4.pdf (accessed 15 May 2023), World Wildlife Fund, Washington, DC, 2006.

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.

Lindsay, J. B.: The Whitebox Geospatial Analysis Tools Project and Open-Access GIS, 2022.

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, W., Li, L., Chen, J., Lin, Q., and Chen, H.: Impacts of land use and land cover change and reforestation on summer rainfall in the Yangtze River basin, Hydrol Earth Syst Sci, 25, https://doi.org/10.5194/hess-25-4531-2021, 2021.

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.

Maryland Department of Environment: Maryland's Phase III Watershed Implementation Plan to Restore Chesapeake Bay by 2025, https://mde.maryland.gov/programs/Water/TMDL/TMDLImplementation/Pages/Phase3WIP.aspx (accessed 15 May 2023), 2019.

Monteith, J. L.: Evaporation and environment., Symp Soc Exp Biol, 19, 205–234, 1965.

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.

Myers, D. T., Jones, D., Oviedo-Vargas, D., Schmit, J. P., Ficklin, D. L., and Zhang, X.: Seasonal landcover variation and environmental modeling data https://doi.org/10.17632/bbb9xbpv22.3 [Dataset]., Mendeley Data, 1, https://doi.org/10.17632/BBB9XBPV22.3, 2022.

Naha, S., Rico-Ramirez, M. A., and Rosolem, R.: Quantifying the impacts of land cover change on hydrological responses in the Mahanadi river basin in India, Hydrol Earth Syst Sci, 25, https://doi.org/10.5194/hess-25-6339-2021, 2021.

Nash, J. E. and Sutcliffe, J. v.: River flow forecasting through conceptual models part I - A discussion of principles, J Hydrol (Amst), 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

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.

Ni, X., Parajuli, P. B., Ouyang, Y., Dash, P., and Siegert, C.: Assessing land use change impact on stream discharge and stream water quality in an agricultural watershed, Catena (Amst), 198, https://doi.org/10.1016/j.catena.2020.105055, 2021.

Nkwasa, A., Chawanda, C. J., Msigwa, A., Komakech, H. C., Verbeiren, B., and van Griensven, A.: How Can We Represent Seasonal Land Use Dynamics in SWAT and SWAT+ Models for African Cultivated Catchments?, Water 2020, Vol. 12, Page 1541, 12, 1541, https://doi.org/10.3390/W12061541, 2020.

Norris, M., Pieper, J., Watts, T., and Cattani, A.: National Capital Region Network Inventory and Monitoring Program Water Chemistry and Quantity Monitoring Protocol Version 2.0 Water chemistry, nutrient dynamics, and surface water dynamics vital signs, Natural Resource Report NPS/NCRN/NRR—2011/423, 2011.

NRCS: Technical Release 55: Urban Hydrology for Small Watersheds, USDA Natural Resource Conservation Service Conservation Engineering Division Technical Release 55, https://tamug-ir.tdl.org/handle/1969.3/24438 (accessed 16 May 2023), https://doi.org/Technical Release 55, 1986.

Pianosi, F. and Wagener, T.: A simple and efficient method for global sensitivity analysis based oncumulative distribution functions, Environmental Modelling and Software, 67, 1–11, https://doi.org/10.1016/j.envsoft.2015.01.004, 2015.

Pianosi, F., Sarrazin, F., and Wagener, T.: A Matlab toolbox for Global Sensitivity Analysis, Environmental Modelling and Software, 70, 80–85, https://doi.org/10.1016/j.envsoft.2015.04.009, 2015.

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.

Ries, K. G., Newson, J. K., Smith, M. J., Guthrie, J. D., Steeves, P. A., Haluska, T., Kolb, K., Thompson, R. F., Santoro, R. D., and Vraga, H. W.: StreamStats, version 4, US Geological Survey, https://doi.org/10.3133/FS20173046, 2017.

Ritchie, J. T.: Model for predicting evaporation from a row crop with incomplete cover, Water Resour Res, 8, 1204–1213, https://doi.org/10.1029/WR008i005p01204, 1972.

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.

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

Sugarbaker, L. J., Constance, E. W., Heidemann, H. K., Jason, A. L., Lukas, V., Saghy, D. L., and Stoker, J. M.: USGS Circular 1399: The 3D Elevation Program Initiative— A Call for Action, https://pubs.usgs.gov/circ/1399/ (accessed 16 May 2023), 2014.

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.

Tong, S. T. Y., Liu, A. J., and Goodrich, J. A.: Assessing the water quality impacts of future land-use changes in an urbanising watershed, Civil Engineering and Environmental Systems, 26, https://doi.org/10.1080/10286600802003393, 2009.

USDA: National Agriculture Imagery Program (NAIP) - Catalog, https://www.usgs.gov/centers/eros/science/usgs-eros-archive-aerial-photography-national-agriculture-imagery-program-naip (accessed 16 May 2023), 2022.

660 USGS: National Water Information System data available on the World Wide Web (USGS Water Data for the Nation), United States Geological Survey, https://waterdata.usgs.gov/nwis/rt (accessed 16 May 2023), 2022.

Utz, R., Bidlack, S., Fisher, B., Kaushal, S., and Ryan Utz, C. M.: Urbanization drives geographically heterogeneous freshwater salinization in the northeastern United States, J Environ Qual, https://doi.org/10.1002/JEQ2.20379, 2022.

Venter, Z. S., Barton, D. N., Chakraborty, T., Simensen, T., and Singh, G.: Global 10 m Land Use Land Cover Datasets: A

665 Comparison of Dynamic World, World Cover and Esri Land Cover, Remote Sensing 2022, Vol. 14, Page 4101, 14, 4101, https://doi.org/10.3390/RS14164101, 2022.

Di Vittorio, A. V., Mao, J., Shi, X., Chini, L., Hurtt, G., and Collins, W. D.: 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.

670 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.

675 Willmott, C. J., Robeson, S. M., and Matsuura, K.: Short Communication A refined index of model performance, International Journal of Climatology, 33, 1053–1056, https://doi.org/10.1002/joc.2419, 2012.

Yang, X., Rode, M., Jomaa, S., Merbach, I., Tetzlaff, D., Soulsby, C., and Borchardt, D.: Functional Multi-Scale Integration of Agricultural Nitrogen-Budgets Into Catchment Water Quality Modeling, Geophys Res Lett, 49, https://doi.org/10.1029/2021GL096833, 2022.

680 Zadeh, F. K., Nossent, J., Sarrazin, F., Pianosi, F., van Griensven, A., Wagener, T., and Bauwens, W.: Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model, Environmental Modelling and Software, 91, 210–222, https://doi.org/10.1016/j.envsoft.2017.02.001, 2017.

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

685 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.

Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., Li, Y., Toman, E., Mallick, B., Zhang, X., and Brown, M.:

690 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.

695   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.