Reviewer 1

Summary

The work uses a calibrated fully-distributed ecohydrological model to explore the nitrogen sources, and the transport and transformation processes within a small exurban catchment. The manuscript seems to contribute to important process understanding, but the current presentation needs substantial revision to legible contribution to existing literature. My key concerns are: 1) lack of clarification on research novelty in the context of existing literature; 2) questionable capacity of the calibrated model to represent key processes, which raise further question on whether the model is appropriate to infer nitrogen dynamics from. I therefore suggest this manuscript to be returned to the authors for substantial revision and resubmission.

Thanks for your helpful comments to our manuscript. We addressed your concerns to our study in novelty and model uncertainty for N dynamics accordingly in responds to your General Comments:

We highlighted our revisions to the manuscript by using blue and calibri font with shading.

- Updates are highlighted in blue
- Responses in times new roman font

General Comments

1. What is the novelty of this study? A clear statement of this is needed with respect to the key knowledge gap in the existing literature: what is missing from current literature, and why are they important to consider. At present, the review of process-based modelling literature seems technically comprehensive, but it does not explain why the current study is needed as a useful addition to literature.

Thanks for the comment. We highlighted the novelty of our study in the abstract and in the introduction and discussion. Our study addresses the **distribution and interaction of hillslope ecohydrological processes** in transporting natural and **human sources of nitrogen** in a long term monitored suburban watershed. Understanding processes and interactions at these scales promotes the design of retention features.

To our knowledge, our model is the first fully distributed hydrologic model that includes i) spatial and temporal human-induced N and water sources at the household level, ii) hillslope ecohydrological processes for routing and cycling water, carbon, and nitrogen. These processes are necessary to identify the space/time distribution of "hot spots" of N retention at scales amenable to restoration.

A significant aspect of the model is that it is calibrated for hydrologic processes restricted to soil and subsurface hydraulic parameters. It is not calibrated for biogeochemical processes which are subject to change with restoration activities. In contrast, the current set of ecohydrological models typically calibrate patch (grid cells, elements) to stream transfer, and biogeochemical cycling parameters.

Abstract

Line 21:

We evaluated how the spatial and temporal distribution of nitrogen sources interacts with ecohydrological transport and transformation processes along surface/subsurface flowpaths to nitrogen cycling, and export. Embedding distributed household sources of nitrogen and water within hillslope hydrologic systems influences the development of both planned and unplanned "hot spots" of nitrogen flux and retention in suburban ecosystems.

Line 29:

With the model is calibrated for subsurface hydraulic parameters only and without calibrating ecosystem and biogeochemical processes, the model predicted mean […]

In the Introduction, we thoroughly reorganized the order of paragraphs and firstly highlighted why understanding ecohydrological processes at "hillslope level" is required for planning Best Management Practices and promote N retention.

Line 47:

BMPs can be both structural (e.g., constructed wetlands) and non-structural (e.g., changing fertilization and irrigation regimes). In addition to planned BMPs, spontaneously developed "hot spots" (Palta et al., 2017) may be responsible for a large share of nutrient retention, and therefore should be identified and protected. Both planned and unplanned retention features exist at very localized, sub-hillslope scales. Therefore, gaining a comprehensive understanding of the hillslope level ecohydrological behaviours and interactions between i) ecosystems and human derived nitrogen sources and ii) flowpath modification can lay the foundation for effectively mitigating these environmental issues through spatially well-conceived and sustainable management practices.

Then, we briefly reviewed how urban water quality is degraded by excessive human-induced N loads, emphasizing the widely used septic systems in suburban areas.\

Line 60:

In the United States, about 20% of households (26.1 million) are reported to be served by septic systems in 2007 (U.S. EPA, 2008). Through our work in Baltimore Ecosystem Study, low density suburban areas have been shown to produce the highest NO_3^- load per unit developed land among different land uses, degrading local and downstream water quality (Groffman et al., 2004; Zhang et al., 2022).

We then discussed the research gap in current semi-distributed models in aspects of incapable of including i) household scale human water and N loads contributing the majority of N inputs in suburban watersheds in distinct landscape positions and ii) hillslope hydrologic flow paths to meet the planning purposes to design BMPs to reduce N export. We also discussed data-driven approaches which could include additional N inputs, but hillslope-level N transport and transformation is still missing.

Line 69:

With rapid suburban and exurban sprawl, decision makers are facing environmental challenges which requires detailed planning for siting BMPs effectively in watersheds to promote N retention, reduce N export in streams, and protect water quality. These include both constructed and "inadvertent" biogeochemical hot spots at specific hillslope locations (e.g., swales, wetlands, riparian areas) on N retention at resolutions required for landscape design. However, commonly used modelling frameworks could not couple distributions and interactions of hillslope ecohydrological processes in transporting and transforming natural and human-induced N sources to understand or predict local (neighbourhood or hillslope) scale N transport and retention. Semi-distributed. Semi […] lack(s) hillslope water and nutrient mixing along interacting surface/subsurface hydrologic flowpaths […]

Line 82:

Data-driven approaches, such as SPARROW (Ator & Garcia, 2016; Smith et al., 1997), are also developed to assess large scale water quality in streams by nonlinear regression from gauged discharge and solute concentrations. However, these models also do not investigate hillslope-scale transport and transformation processes. In addition, there does not exist the data at hillslope scales to develop sufficient data-based approaches to understand and predict retention processes (e.g., denitrification, uptake, immobilization).

Then, we emphasized, though fully distributed hydrologic models, such as MIKE-SHE, could simulate hillslope hydrology and biogeochemistry, they currently have no modules to include the household-level N inputs developed.

Line 87:

Fully distributed hydrology models, such as MIKE-SHE (Abbott et al., 1986a, 1986b) and RTM-PiHM (Bao et al., 2017; Zhi et al., 2022), ParFlow (Maxwell, 2013) and RHESSys (Tague & Band, 2004) could explicitly couple hillslope hydrologic and biogeochemical processes that are required to understand transport and transformation of these human-induced N loads along hydrologic flowpaths from upland to stream.

Lastly, we wanted to highlight that our model is designed to be generalized to watersheds without long-term water chemistry observations which are quite expensive to acquire. In other words, we do not calibrate our parameters for N inputs (e.g., fertilization and septic loads) or processes but only soil hydraulics against streamflow records. If the model could reasonably estimate NO_3^-, it compromises the generalization of the model.

Line 102:
Lastly, the framework should be capable to be extrapolated to watersheds without water chemistry data which are less available than discharge records worldwide. It would be a valuable feature of the framework to estimate nutrient dynamics reasonably if calibrating only hydrologic parameters could provide reasonable estimation of N dynamics. Calibrating nutrient dynamics may not allow generalization to watersheds without chemistry records or extrapolation to conditions in which water quality BMPs are implemented.

2. The Introduction started discussion different types of models and their pros/cons from an earlier stage, lots of them are about inclusion of key processes (e.g., L55: hillslope water and nutrient mixing along hydrologic flowpaths). But for the readers' benefit, it might be clearer by adding a separate paragraph before introducing all the models, to discuss the theory about key processes at the particular spatial/temporal scale that you are interested in? Then you can start discussing and contrasting models based on their process representation.

Thanks for this suggestion. As in the response to Comment 1, we thoroughly reorganized the Introduction to improve its flow and readability. After the opening paragraph, we firstly emphasized the urgency to understand how excessive human N inputs affect water quality in urban watersheds, and then discussed the research gaps in current frameworks by comparing the semi- and fully distributed models and their limitations.

Lastly, we highlighted that our RHESSys model could be augmented to fill these research gaps in other models and advance our understanding to N dynamics of urban watersheds while recognizing some of the scale (watershed size) limitations.

Line 107:

The Regional Hydro-Ecological Simulator System (RHESSys, Tague & Band, 2004) is designed to meet all requirements for the framework, which is an ecohydrological model that simulates mass balances of water, C, and N of a watershed including hydrologic and biogeochemical stores and cycling. […]. In this study, we augmented RHESSys to include household-level transfer of groundwater for lawn irrigation and domestic water use, with domestic water use routed to septic spreading fields. With coupling hillslope hydrology and biogeochemistry at spatially connected patches, RHESSys could estimate spatiotemporal patterns of […] in spatially explicit manners. In summary, by adding modules of lawn irrigation, fertilization, and septic releases (see Sect. 2.3) that are commonly found in suburban areas, RHESSys is designed with the capacity to simulate the comprehensive ecosystem dynamics and feedbacks of introduced spatially explicit lawn irrigation, fertilization, and septic releases that are commonly found in suburban areas, at resolutions commensurate with human management of the landscape. This facilitates scientific assessment of small-scale human activity and modification to land cover and infrastructure in expanding suburban and exurban areas.

3. You have a comprehensive review of process-based water quality models, what about the data-driven ones? The latter seem very useful to explain processes/changes at larger scales (e.g.,) – what's their relevance to your study? I think this comment can be potentially addressed once you have resolved my Comment #2.

   Thanks for the comment. We addressed this in our response to Comment 1. Our model, compared to data-driven water quality models, is capable of providing the comprehensive representation of overall N cycling inside the watershed, which includes interacting processes (e.g., denitrification, uptake, nitrification, etc.) beyond NO3- concentration at the outlet. Data-driven water quality models (e.g., SPARROW) may capture the change of stream N concentrations and loads due to land cover changes from urbanization within a watershed, but is not designed to estimate the impacts of small scale (below the level of a catchment) inputs of water and nitrogen and response to retention features. The data driven methods are useful for estimation of large-scale loads and concentrations of stream network N, but data to develop methods at the landscape scale we address are lacking. Therefore, we added a few sentences from line 108 to 112 contrast both approaches:

   Line 82:
   Data-driven approaches, such as SPARROW (Ator & Garcia, 2016; Smith et al., 1997), are also developed to assess large scale water quality in streams by nonlinear regression from gauged discharge and solute concentrations. However, these models also do not investigate hillslope-scale transport and transformation processes. In addition, there does not exist the data at hillslope scales to develop sufficient data-based approaches to understand and predict retention processes (e.g., denitrification, uptake, immobilization).

   References

   Ator, S. W., & Garcia, A. M. (2016). Application of SPARROW modeling to understanding contaminant fate and transport from uplands to streams. *Journal of the American Water Resources Association, 52*(3), 685-704. https://doi.org/10.1111/1752-1688.12419

   Smith, R. A., Schwarz, G. E., & Alexander, R. B. (1997). Regional interpretation of water-quality monitoring data. *Water resources research*, *33*(12), 2781-2798. https://doi.org/10.1029/97WR02171

4. The Methods section states that for model calibration 'the parameter set yielding the highest NSE was used to simulate ecohydrological processes' – this does not allow for structural uncertainty, is there any implication on your results? It might be a more robust practice to include multiple sets of 'better performing' parameters and then compare how they represent the hydrology; the current calibrated model seems to capture broad seasonality patterns, but either misses a few high-flow events or is a bit delayed compared to the observation (Figure

3), but it's difficult to tell as the lines for observations and simulations in Figure 3 are on top of each other – it would be clearer to use dots and lines in showing the two sets of data

Thanks for the comment. Firstly, our parameters were calibrated against the streamflow observation only, which is provided by USGS at daily scale. To quantify the uncertainty of model simulations, we **performed another round of calibration in water year 2013 to 2015, with validation period of water 2016 to 2017.** We chose **50 behavioral parameter sets yielding the NSE values ranging from 0.5 to 0.69 in the calibration period. All these parameter sets were restricted to have the gw2 (% groundwater loss to stream, Table 1) lower than 0.5 to avoid simulating too flashy groundwater dynamics.** We found these parameter sets all yield similar hydrologic behaviors, and the uncertainty boundary of NO_3^- reasonably captured the majority of our observations, despite that we do not calibrate any N-related parameters.

We repeat that the goal of calibrating hydrologic parameters (subsurface hydraulic parameters) only, was to avoid calibrating N cycling dynamics which may compromise the generalization of the model.

Line 205:

We set the calibration period from water year 2013 to 2015 and validation period from water year 2016 to 2017. The original parameter values derived from SSURGO were further calibrated by multipliers to vary their magnitudes but preserve the spatial patterns of soil hydraulic properties (Fig. A2). Specifically, the simulated streamflow was used to calibrate against the daily USGS discharge records (Gage ID: 01583580). From four thousands of parameter set realizations randomly chosen within specified limits, behavioural sets are chosen as yielding Nash-Sutcliffe efficiency (NSE; Nash & Sutcliffe, 1970) greater than 0.5 and fraction of groundwater loss to stream (i.e., gw_2 in Table 1) less than 0.5 to estimate the ensemble means and uncertainties of model simulations. The latter condition was enforced to regulate the flashiness of groundwater dynamics, as BARN is found to have large saprolite storage to provide steady baseflow (Putnam, 2018). To assess uncertainty, we reported the 95% uncertainty boundaries for simulated streamflow and NO_3^- concentration and load from. Lastly, we noted that no calibration was performed for N inputs (e.g., fertilization rate and septic load) or N cycling/transport processes in the model, as an important aim of our methods is to evaluate the capacity of our model to regionalize to watersheds where no water chemistry but only streamflow observations were available.

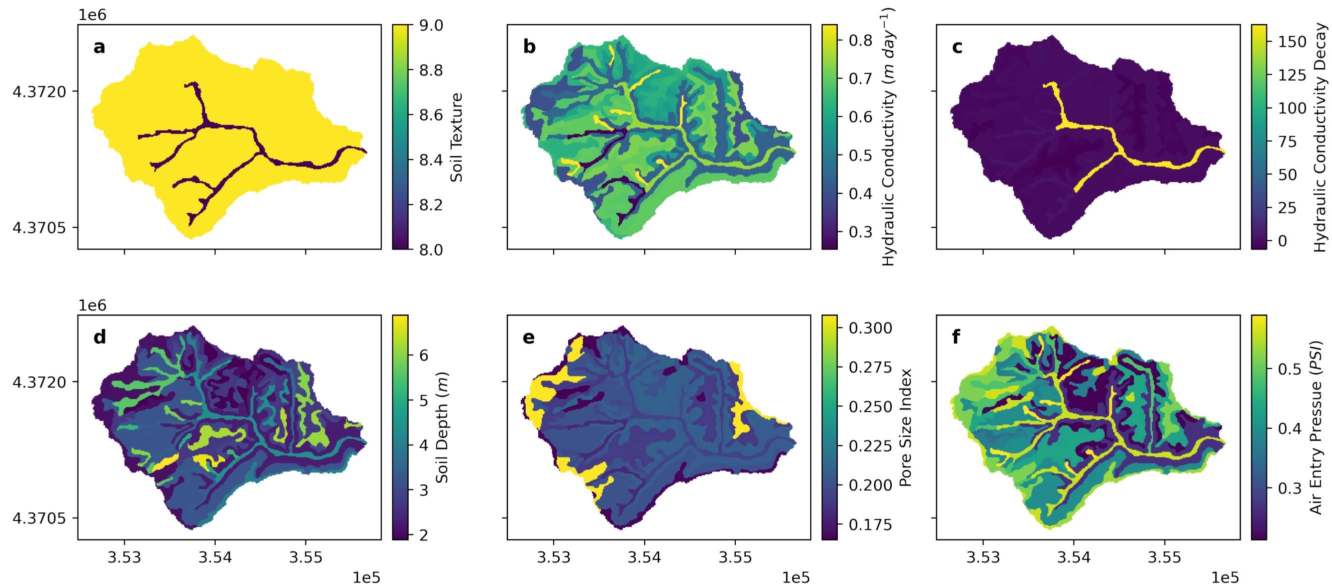


Figure A2. SSURGO (USDA, 2019) derived (a) soil texture, (b) lateral and vertical saturated hydraulic conductivities at surface (m day^{-1}), (c) lateral and vertical decay rates for lateral and vertical hydraulic conductivities, (d) soil depth (m), (e) pore size index, and (f) air entry pressure (pounds inch^{-2}) for Baisman Run.

We then found 50 behavioral parameter sets meeting the requirements. We were also able to quantify the uncertainty of our model from these behavioral simulations.

Streamflow uncertainty (Line 340):

The range of calibrated multipliers are listed in Table 1, and the distributions are shown in Fig. A3. In the calibration period (i.e., water year 2013 to 2015, Fig. 3a), the ensemble of simulated mean (standard deviation) daily streamflow was 1.24 (□0.03) mm day^{-1}, with NSE of 0.63 (between 0.5 and 0.69) compared to the USGS observed 1.38 mm day^{-1}. In the validation period (Fig. 3b), the simulated ensemble mean (standard deviation) streamflow was 0.91 (□0.03) mm day^{-1}, with NSE of 0.58 (between 0.44 to 0.64) compared to the USGS's 0.86 mm day^{-1}.

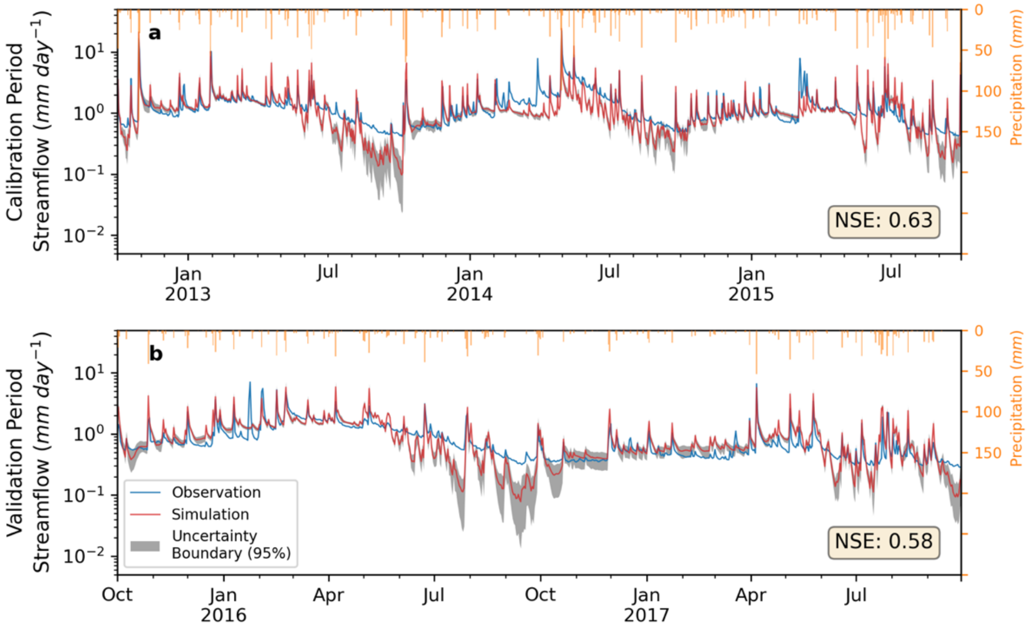


Figure 1. The ensemble mean of daily streamflow from simulations (red) with NSE greater than 0.5 and USGS observations (blue), with the daily 95% uncertainty range from 50 simulations in grey for the (a) calibration (Oct. 2012 – Sep. 2015) and (b) validation (Oct. 2015 – Sep. 2017) period. All simulations turned on irrigation, lawn fertilization, and septic processes

We note that we modified our Figure 3 to better contrast of the two lines. We added the 95% uncertainty range to the streamflow plot. Considering our data are at daily scale, plotting in dots would still have a lot of overlap and may be noisier than the line plot. To help readers to contrast the two lines better, we made the lines thinner and increased the transparency of our simulation line so both lines can be detected.

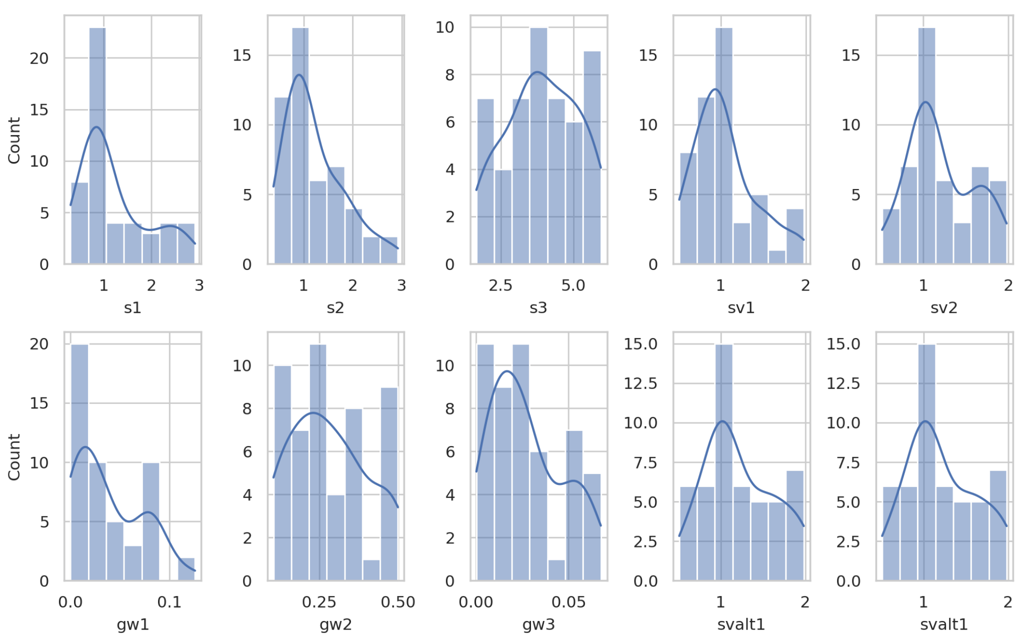


Figure A3. Distributions of multipliers to RHESSys parameters based on 50 calibrated behavioral parameter sets.

NO_3^- Concentration (Line 366):

We calculated weekly means of NO_3^- load and concentration of behavioural simulations. In our 5-year study period, the ensemble mean NO_3^- concentrations (Fig. 4a) for scenarios *none*, *septic only, fertilization only, and both* were 0.34, 0.77, 0.87, and 1.43 mg NO_3^--N L^{-1}, respectively (Table 4). The mean long-term observed concentration at the BARN USGS gauge was 1.6 mg NO_3^--N L^{-1}. Thus, the simulated bias of mean NO_3^- concentration considering both fertilization and septic loads decreased significantly from -1.26 mg NO_3^--N L^{-1} in the scenario *none* to 0.17 mg NO_3^--N L^{-1} in the scenario *both*. The 95% uncertainty boundary of weekly NO_3^- concentration in scenario *both* captured 67% of the weekly sampled observations.

Load (Line 377):

The in-stream NO_3^- load (Fig. 4b) followed a similar trend as concentration, and the bias was reduced substantially from scenario *none* to *both* when fertilizer and septic loads were included. Scenario *none* underestimated NO_3^- load by 6 (-81%) kg NO_3^--N ha^{-1} $year^{-1}$, and the scenario *both* decreased the bias substantially to -0.77 (-10%) kg NO_3^--N ha^{-1} $year^{-1}$. The seasonality was also well simulated by our model. The ensemble mean loads (Table 3) in fall and winter were accurately captured with close-t- zero bias compared to the observations, and the bias in spring and summer was slightly higher.

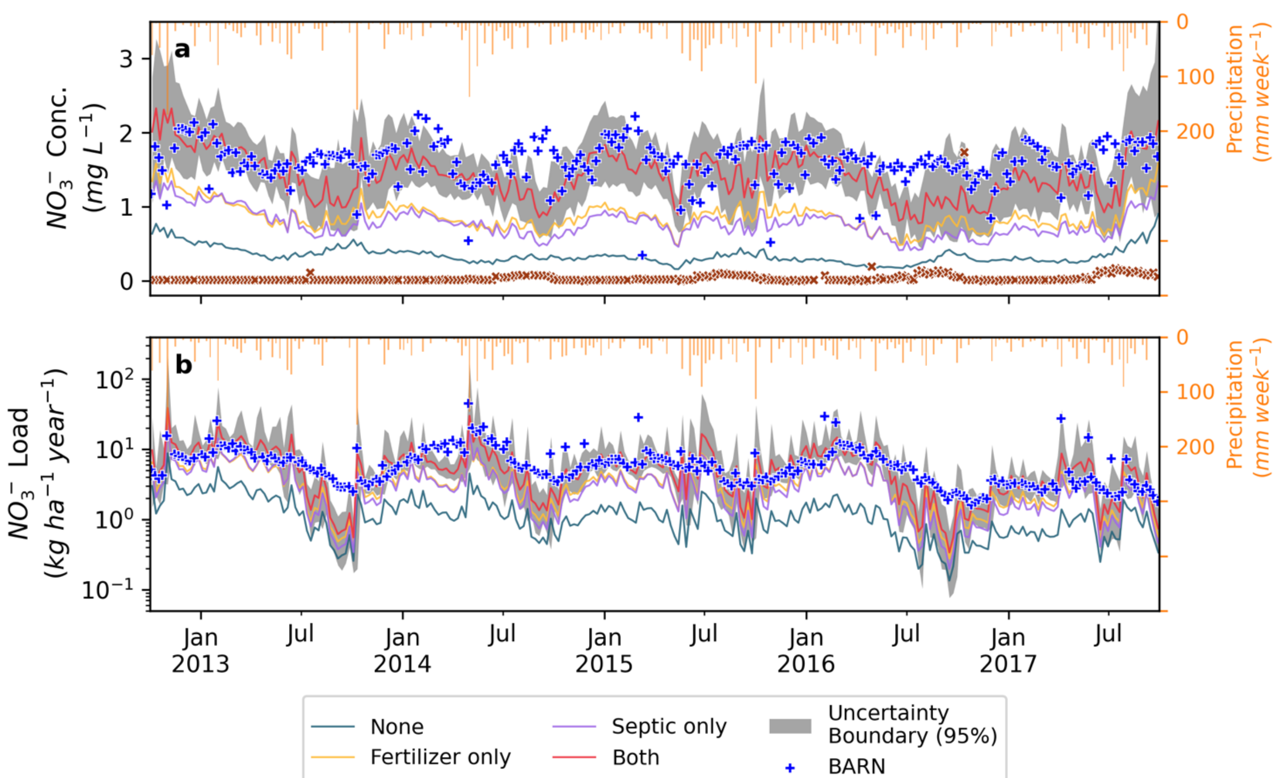


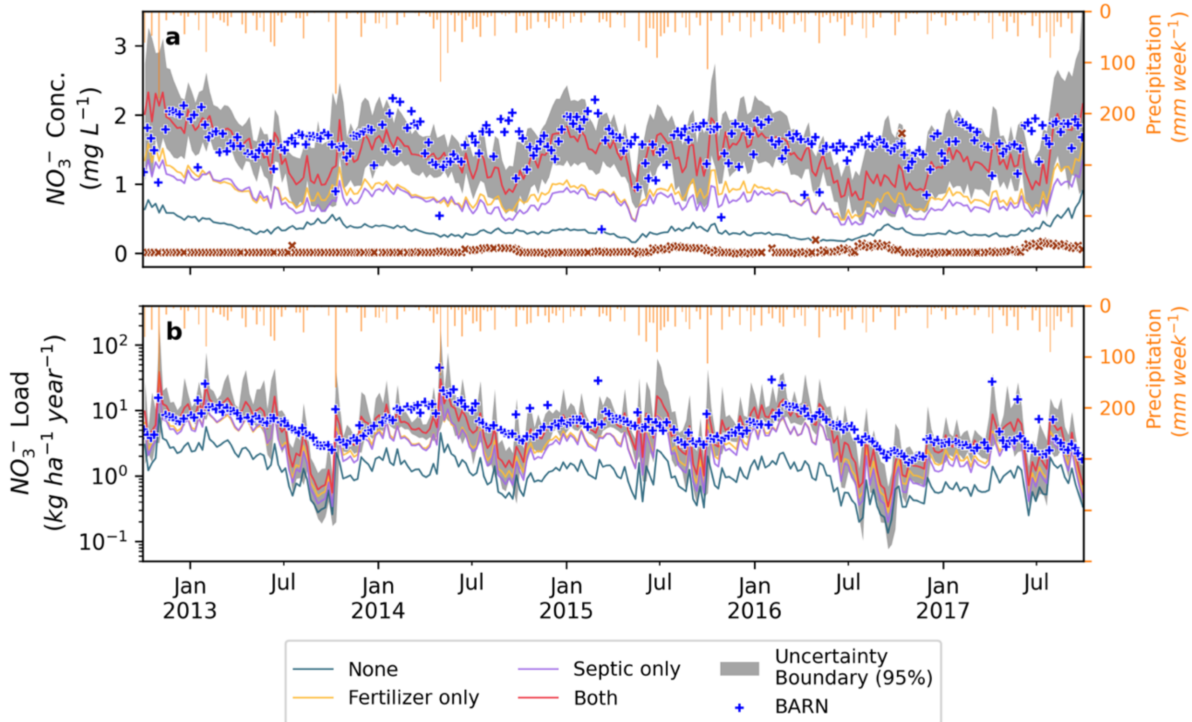
Figure 4. Ensemble weekly mean (a) NO_3^- concentration and (b) load at the outlet of Baisman Run over the entire study period (water year 2013 to 2017). The 95% uncertainty boundary for scenario *both* was shown in grey.

Table 1. Mean weekly NO_3^- concentration (mg N L^{-1}) and load (kg N ha^{-1} $year^{-1}$) from calibrated simulations for BES weekly observations (BARN and POBR) and RHESSys simulation scenarios in each season and the entire study period from water year 2013 to 2017. Standard deviations from behavioural simulations for all scenarios were included below the mean values.

Variables	Season	Observation		RHESSys Scenarios			
		BARN	POBR	Both	Septic Only	Fertilizer Only	None
Concentration (mg N L^{-1})	Spring	1.5	0.02	1.4 (±0.12)	0.76 (±0.08)	0.77 (±0.05)	0.27 (±0.03)
	Summer	1.6	0.07	1.26 (±0.13)	0.68 (±0.1)	0.79 (±0.1)	0.33 (±0.06)
	Fall	1.57	0.06	1.41 (±0.23)	0.77 (±0.15)	0.94 (±0.17)	0.41 (±0.09)
	Winter	1.75	0.01	1.63 (±0.18)	0.88 (±0.12)	0.96 (±0.1)	0.35 (±0.05)
	Mean	1.6	0.04	1.43 (±0.16)	0.77 (±0.11)	0.87 (±0.1)	0.34 (±0.06)
Load (kg ha^{-1} $year^{-1}$)	Spring	10.93	0.01	8.86 (±0.63)	4.84 (±0.42)	4.77 (±0.31)	1.62 (±0.16)
	Summer	5.88	0.02	4.72 (±0.36)	2.49 (±0.25)	2.81 (±0.23)	1.06 (±0.16)
	Fall	4.72	0.01	4.72 (±0.39)	2.57 (±0.26)	3 (±0.27)	1.23 (±0.16)
	Winter	8.38	0.01	8.42 (±0.68)	4.61 (±0.46)	4.91 (±0.38)	1.81 (±0.18)
	Mean	7.44	0.01	6.68 (±0.47)	3.63 (±0.33)	3.87 (±0.27)	1.44 (±0.16)

5. I think the abovementioned issue in simulating hydrology also brings question on whether the water quality dynamics are well represented by the model. Besides a consistent lower bias (i.e., for ‘both’ scenario has an approx. -50% average bias, Figure 5), the simulated seasonality of NO3 concentrations also seem to differ from the observation too. I’m not convinced that this calibrate model is reasonable to further infer on hydrological/water quality processes. Has any model performance metric been calculated for NO3?

Thanks for the comment. Our apology for put an incorrect figure for Figure 5, which used the wrong low fertilization inputs values from Law et al. (2004) due to my coding mistakes. Except for this figure, all other results were reported using the correct fertilization rates. We have corrected this figure as below.



We discussed the details in Discussion that there are uncertainties in hydrologic behaviors and parameterization which could affect the simulation of NO_3^- concentration, especially during the end of growing season (Fig. 3) when uncertainty of water usage and vegetation behaviors are not fully understood. Also, the spatial and temporal patterns of N inputs were assumed uniform for all households in the watershed, but the variations could significantly affect the N transport and transformation in the watershed. We also note that our observation samples were all collected under non-storm conditions, which could be quite different from our simulations which include all weather conditions. In summary, without calibrating N-related parameters of RHESSys, our model yield quite reasonable NO3- concentration compared to the observed records.

Line 483:

Considering that no N-related parameters were calibrated, the reasonable NO_3^- simulations suggest the model can provide sufficient assessment of the effects of household water and nutrient management on N transport, transform, and export in suburban watersheds when only discharge but no NO_3^- observations are available. The uncalibrated parameters of vegetation and domestic water usage introduced uncertainty in hydrologic and biogeochemical processes of our model, which may cause bias in streamflow and N cycling especially in the dry periods during the growing season. In these periods, our model might retain excessive N in the upland through denitrification and uptake, leaving little transported to streams. In addition, we assumed identical N inputs for all households in BARN, but the actual fertilization and septic effluents may have considerable spatial, and temporal variations which could impact the N cycling and transport significantly. Specifically, we used the annual fertilization rate on lawns as 84 kg N ha^{-1} from Law et al. (2004) in which the reported range of annual fertilization was from 10.5 to 369.7 kg N ha^{-1}. [...] Lastly, we noted that the observations of weekly NO_3^- from BES were collected in conditions without large storm flows, but our model simulated NO_3^- under various weather conditions. Bias between our model simulation and the observations is unavoidably expected.

6. There are some key information lacking in the Methods, some examples are listed below but they highlight need for a substantial improvement of the Methods section:

- Section 2.2 on calibration, was the model calibrated to only the streamflow record or with the water chemistry concentration data as well, and at which gauge? Please specify.

Thanks for the comment. We highlighted in the responses to previous Comments that our calibration was performed only against the daily USGS discharge records, and no N-related parameters were calibrated. We added the USGS gage ID (Gage ID: 01583580) at line 216.

Line 221:

Lastly, we noted that no calibration was performed for N inputs (e.g., fertilization rate and septic load) or N cycling/transport processes in the model, as an important aim of our methods is to evaluate the capacity of our model to regionalize to watersheds where no water chemistry but only streamflow observations were available.

- In Table 1, what does the column 'sensitivity parameter' refers to? Also, for completeness, the table should also present the original parameter values estimated from SSURGO soils dataset besides the calibrated multipliers.

Thanks for this comment. We included physical meanings of parameters in Table 1 and the original SSURGO values in Fig B2. The SSURGO values were estimated for each type of soils and varies among patches, therefore we could not include a single value for each parameter but showed the maps of these values in Fig. B2.

We added a sentence for readers to check supplementary for more information about SSURGO soil at line 202.

[...] we calibrated eight parameters (Table 1) for subsurface properties (i.e., lateral and vertical saturated hydraulic conductivities and their decay rates, pore size index, and air entry pressure) with initial estimates (Fig. A2) from the SSURGO soils dataset (USDA, 2019) and deeper groundwater processes (i.e., bypass seepage from surface and shallow saturated soil, and drainage rate to stream). [...]

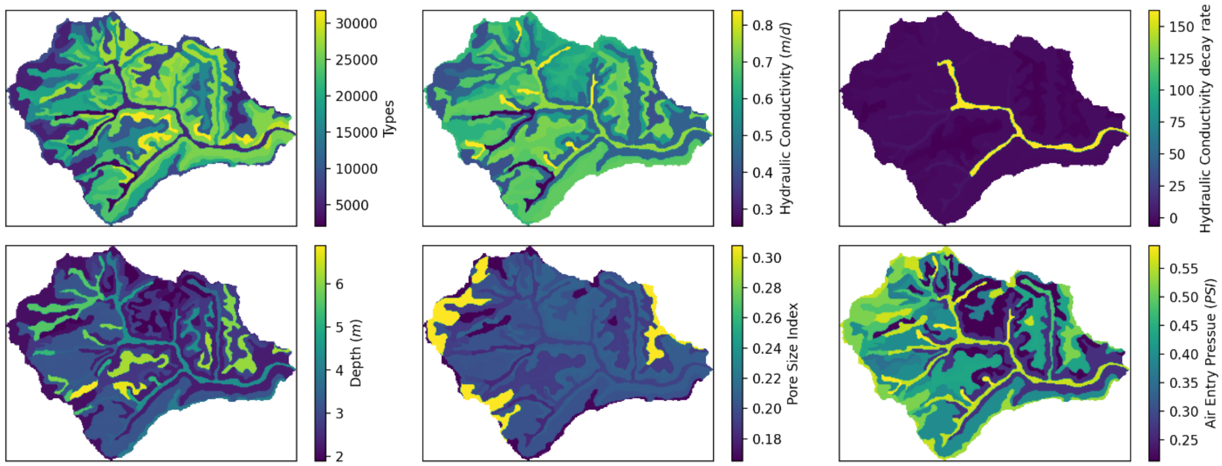


Figure A2. Soil properties derived from SSURGO dataset. These values are calibrated against USGS observations and modified by multipliers listed in Table 1.

Table 2. RHESSys parameters being calibrated and their physics (Tague and Band, 2004). Calibrated results shown as ranges of multipliers to original soil properties (Fig. A2 & A3) and groundwater component generating behavioural simulations with NSE greater than 0.5 for streamflow.

Parameter Groups	RHESSys Parameter Abbreviations		Detail	Source	Unit	Multiplier Range
Lateral soil hydraulics	s	m_l	Decay rate of lateral saturated hydraulic conductivity with depth	USDA SSURGO, 2019	-	0.31 – 2.91
		K_{sat0_l}	Lateral saturated hydraulic conductivity at the soil surface		m day^{-1}	0.38 – 2.93
		z	Soil depth		m	1.65 – 5.95
Vertical soil hydraulics	sv	m_v	Decay rate of vertical saturated hydraulic conductivity with depth	USDA SSURGO, 2019	-	0.51 – 1.98
		K_{sat0_v}	Vertical saturated hydraulic conductivity at the soil surface		m day^{-1}	0.52 – 1.98
Soil properties	svalt	b	Pore size index	USDA SSURGO, 2019	-	0.51 – 1.98
		φ_{ae}	Air entry pressure		pounds inch^{-2}	0.5 – 1.05
Groundwater dynamics	gw	gw_1	Fraction of bypass from the saturated zone to groundwater storage		-	0 – 0.13
		gw_2	Fraction of loss from groundwater storage to stream		-	0.03 – 0.5
		gw_3	Fraction loss from surface to groundwater storage		-	0 – 0.07

- How are rainfall routing and runoff handled by the model? Are there any parameter to calibrated related to the rainfall-runoff processes?

The rainfall-runoff processes of RHESSys are discussed in detail in Tague and Band (2004). At patch level, rainfall is intercepted by vegetation and infiltrated into its soil layers. Surface and subsurface water is then routed to surrounding patches following hydraulic gradients. In subsurface, water is dynamically routed following gradients between water table elevations. Soil parameters, especially lateral and vertical soil hydraulic conductivity (i.e., **s** and **sv** in Table 1), affect the rainfall-runoff and drainage processes directly and are thus calibrated against the runoff observations. The multipliers will only alter the magnitudes of original SSURGO derived values (Fig. A2) but their spatial patterns are preserved. Soil hydraulic conductivities are assumed to decay exponentially, and the lateral and vertical decay rates (i.e., **m** in Table 1) are also calibrated to regulate water routing in this study. Surface routing features, including road and roof drainages, are also considered, as in Smith et al. (2022).

These parameters are commonly calibrated in previous RHESSys studies, and the routing procedure is detailed in Lin et al. (2021). The routing procedure of RHESSys is complex and well tested in previous studies (Smith et al., 2022). Therefore, to keep the focus of this study on N dynamics, we do not include the routing details in the Method, but provide the reference for readers to check at line 286:

RHESSys requires several subsurface hydraulic parameters to simulate lateral and vertical water flows and route subsurface lateral flow that are calibrated following the procedure detailed in Smith et al. (2022).

References

Smith, J. D., Lin, L., Quinn, J. D., & Band, L. E. (2022). Guidance on evaluating parametric model uncertainty at decision-relevant scales. *Hydrology and Earth System Sciences*, *26*(9), 2519-2539. https://doi.org/10.5194/hess-26-2519-2022

- Figure 2: why is rainfall not considered as a key process? How possible is lawn only irrigated by groundwater but not rain water?

Rainfall is the most important water input of the watershed, and it included to all hydrological processes in RHESSys. The Fig. 2, however, is to highlight the new procedures of our augmentations for hillslope **groundwater redistribution** via. irrigation and septic systems, and these pumped waters were distributed to detention storage first and then follow the original RHESSys hydrological processes. Irrigation amount is regulated by the water stress in Equation 3.

Line 251:

Figure 2. Groundwater extraction for irrigation and septic systems in the RHESSys model. The source water (green arrow) is extracted from groundwater storage of drain-in patches (i.e., house centroids) and redistributed (orange arrow) to surface detention in downstream lawn patches for septic effluents and irrigated lawn patches of a household. After redistribution of source water, infiltration to soil and percolation to hillslope groundwater (yellow arrows) would follow the original processing of RHESSys

- Equation 3: PET and ET – how are there estimates?

Potential evapotranspiration (PET) are estimated using the Penman-Monteith equation (Monteith, 1965) assuming no soil water limitations. PET representing the maximal ET rate at given current meteorological information and land cover, and actual ET is estimated when the rate is regulated by soil moisture level and stomatal conductivity in each patch of our model. When water is not limited, PET and ET could be quite close; During droughts, PET could be much higher than the actual ET due to the low soil moisture level.

We provided the references to help reader refer for procedures and equations the RHESSys model uses to estimate PET and ET:

Line 286:

where PET and ET (mm) represent patch level potential and actual ET, which are estimated daily in RHESSys based on the Penman-Monteith equation (Monteith, 1965) and procedures in Section 5.6 in Tague and Band (2004).

7. The Results section presents a lot of information but there is no direct link of them to the modelling outputs. I think the Methods section misses a sub-section at the end on which model outputs are analysed and how, to answer which research question (which links to the Introduction). This would be very helpful for readers to link the Results section with the rest of the paper.

Thanks for the suggestion. We have first presented results in the Results section, but link those results to research questions in the Discussion section. We added a short paragraph in the end of our Methods Section 2.4 (line 327) to help readers refer to the corresponding sections in Results.

In the Results section, we presented model calibration results in Section 3.1, in-stream NO_3^- dynamics of scenarios in Section 3.2, and ecohydrological changes and N hot spots in Section 3.3, accordingly.

Specific Comments

1. Line 21 – the statement seems too long and might be confusing, can you break this into two sentences, or use labels e.g., i), ii) if a single sentence is used?

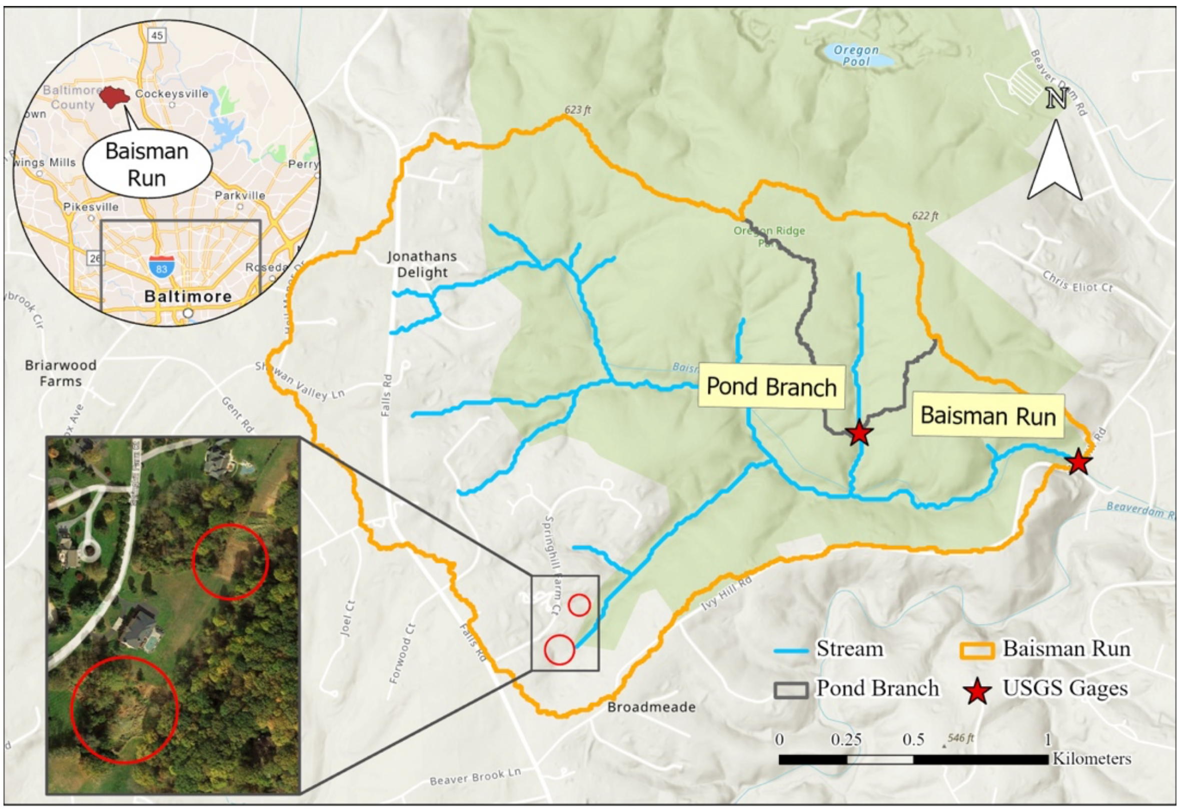
Thanks for the comment. We broke down the sentence to align with other revisions, and changed the original sentence to:

Line 20:

We evaluated how the spatial and temporal distribution of nitrogen sources interacts with ecohydrological transport and transformation processes along surface/subsurface flowpaths to nitrogen cycling, and export. Embedding distributed household sources of nitrogen and water within hillslope hydrologic systems influences the development of both planned and unplanned "hot spots" of nitrogen flux and retention in suburban ecosystems.

2. Figure 1 can be improved by including more information on the study area, including: locations of the two monitoring sites mentioned (01583580, 01583570) and the boundary of the sub-catchment, Pond Branch. The base map would be more informative presented as a map of key land uses (e.g., forest, urban, exurban) instead of a satellite image – it is a bit hard to visualize the land use components from the latter.

Thanks for the comment. I have added the USGS gages and the boundary of Pond Branch in the map as below.



We agreed the satellite image is a noisy background, and we replaced it with a general topographic map with hillshades outlined. The land use map contains 12 classes and could be too noisy for readers to view in the main manuscript, but we also added the land use map in the Appendix as Fig. A1 for readers who want to check the details of the watershed.

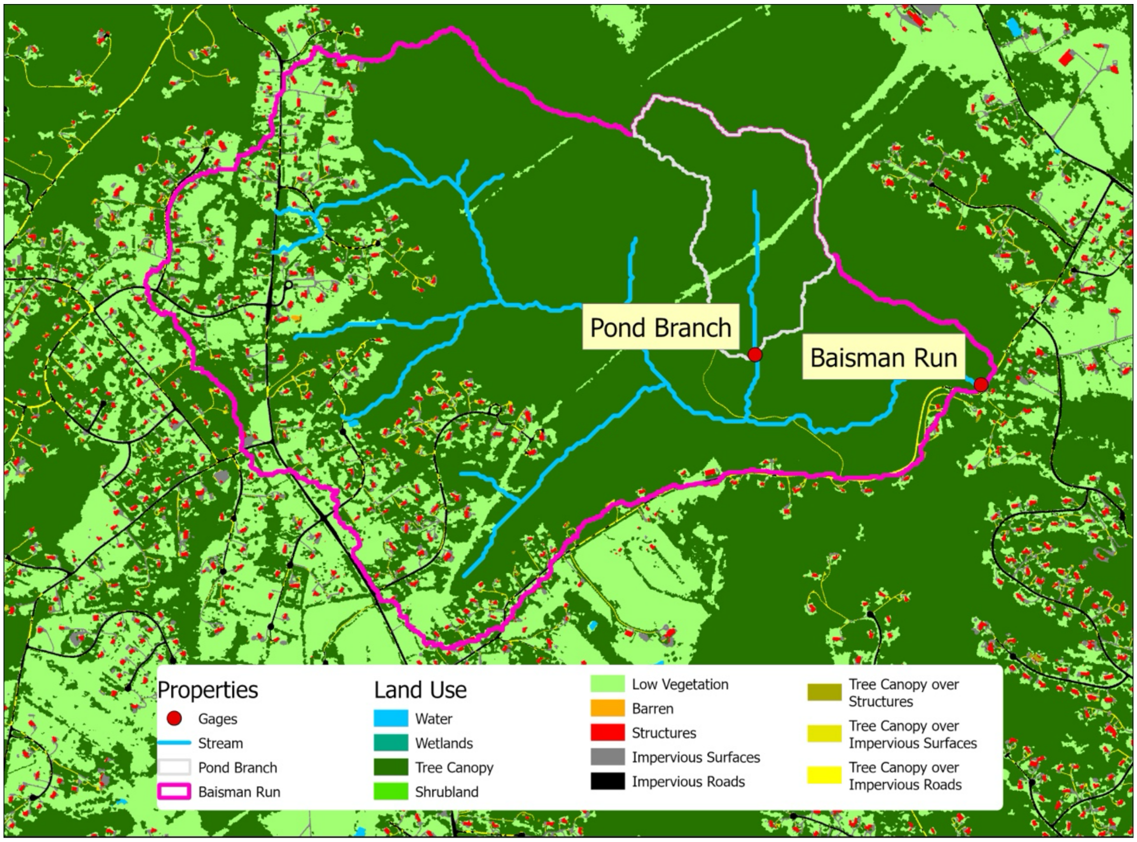


Figure A1. 1-m land use and land cover in Baisman Run from the Chesapeake Bay Conservancy.

3. Relating to my Comment #5, I'm also confused by your statement in L403 'our model underestimated the mean in-stream NO3 – concentration by 0.1 mg NO3 – -N/L (-7%) with stronger variability (Fig. 5)'. In Fig. 5, I see an approx. -50% bias comparing the simulated concentration for the 'both' scenario compared with the observation.

Thanks for the comment. Please refer to our response to your Comment 4 and 5 for the details about the bias in the simulated NO_3^-.