Respond Letter

Dear editor,

We appreciate your time to handle our manuscript. We were able to improve the quality of our manuscript with constructive comments from you and all referees. In this round of revision, we addressed all major and minor comments from all three referees. We also did some minor edits to clarify some concepts and terms, which we think significantly improved the readability and quality of this study. However, we found Referee #3 (report #2) was reviewing the initial submission, not the revised manuscript. Many of the referee’s comments were addressed in the previous revised manuscript, and we carefully respond to these comments by citing the revisions we did in the previous and this round. We wanted to let you know this and would be happy to address any other comments of Referee #3 if she/he has additional questions to our newest version of manuscript.

Please don’t hesitate to let us know about any questions/issues you have regarding to the manuscript. We appreciate your efforts and time spending with our manuscript again.

Best regards,
Ruoyu Zhang (on behalf of all co-authors)

Report #1 (Referee #2)

The revised version of the manuscript entitled "Simulation of spatially distributed sources, transport, and transformation of nitrogen from fertilization and septic system in a suburban watershed" presents an augmented version of the RHESSys Model for simulating the impact of spatially distributed anthropogenic N and water inputs in suburban watersheds. The authors chose a well-monitored watershed to calibrate the augmented ecohydrologic model for subsurface hydraulic parameters and compare the model results against observed NO3- concentrations in streams. The augmented modules of the model rely on a rich composition of data sets obtained in previous surveys and studies. The results prove that the presented approach successfully improves the prediction of nutrient loads in streams by integrating spatially explicit anthropogenic N inputs. It delivers valuable insights about the impact of N retention hot spots and could help planners decide on effective sizing and situating of best management practices. As the presented framework is transferable to simulate the ecohydrology of other watersheds with scarcer or absent nutrient load data, it is a valuable contribution to watershed restoration efforts.

I suggest accepting the manuscript for publication with minor revisions.

Specific comments

#1) The overall quality of the manuscript and presented Figures and results improved significantly. The current state of knowledge and the novelty of the presented approach have been well elaborated. The methodology is now presented clear and well sorted, highlighting the rich input of collected data that backs up the authors augmentation scenarios.
#2) The length of several sentences (some are 3 to 4 lines long) should be revised.
#3) The conclusion should be revised for redundant points.
#4) Typos should be improved (see technical corrections in the report) #5) The contribution of one author is not described in the respective section. Please make sure all authors contributions are described in concordance with the CRediT contributor roles taxonomy (suggested by HESS)

Response:

We appreciate the referee's thorough comments to our revised manuscript. It was your constructive suggestions that help to improve our manuscript significantly. We have carefully addressed all comments below. For your reference, all our responses are in blue color. We also quote and shade texts in the manuscript, and use red color to highlight new changes we made. We also carefully revised the entire manuscript to ensure that there is no grammar mistake found.

l. 28 space/time ?

Thanks for the comment. We have removed "space/time" in the sentence.

l. 29 With?

Thanks for the comment. We rephrased the sentence as:
"With calibrating subsurface hydraulic parameters only and without calibrating ecosystem and biogeochemical processes, the model … "

l. 69 typo: requires

Thanks for the comment. We corrected this as:
"…, decision makers are facing environmental challenges which require detailed planning for siting BMPs effectively in watersheds …"

l. 85 please improve the grammar

Thanks for the comment. We rephrased the sentence as:
"However, these models lack hillslope water and nutrient mixing […]. These interactions are important to simulate the formation of biogeochemical hot spots where potential uptake and retention of nutrients are high."

l. 91 typo soils

Thanks for the comment. We correct "soils" to "soil".

l. 120 add the: […] facilitates [the] scientific assessment […]

Thanks for the comment. We added "the" into the sentence.

l. 223 might "we noted" at this line be a remnant of the reply to the review?

Thanks for the comment. We wanted to highlight the fact that we did not calibrate N-related parameters in this study. We changed to "we emphasized that […]" at line 223.

l. 224 Please correct the typo: from.

Thanks for the comment. We removed "from" in the sentence.

Lastly The standard deviations in the text are not showing the +- symbol right (e.g. l 341, 343 etc.)

Thanks for the comment. The ± symbol was not shown correctly. We have fixed the issue at line 348.

l. 350 the unit mm must should in front of parenthesis

Thanks for the comment. We have moved the unit, including in other sentences throughout the manuscript, in front of the parenthesis.

l. 379 please correct typo close-t- zero

Thanks for the comment. We fixed this as "close-to-zero".

l. 380 simulated instead of simulation

Thanks for the comment. We changed to "simulated".

l. 409 why two values? "(-0.1, -0.3%) "

Thanks for the comment. We added the unit, mm month^{-1}, to avoid the confusion of the bias and percentage of change. We also fixed this issue in the rest of the manuscript.

Line 412:
"ET in lawn patches and septic drainage fields increased to (by) 42.3 (+0.4 mm month^{-1}, 1.0%) and 40.8 (+6.5 mm month^{-1}, 18.9%) mm month^{-}."

l. 431: please improve this sentence: Septic drainage patches (i.e., scenario septic only) was almost 5fold higher (+368%) than the reference scenario none.

Thanks for the comment. We rephrased the sentence in Line 437 as:
"Denitrification rates in septic drainage patches was increased by 368% in scenarios *septic only* compared to in the reference scenario *none* where these patches do not receive additional water and N inputs."

l. 211 Caption Table 1. what does physics of the parameters refer to?

Thanks for the comment. We further clarified this as "physical representations" of calibrated RHESSys parameters in Line 211.

Caption Figure 3: making 2 sentences will help the reader following on the complexity of the figure.

Thanks for the suggestion. We rephrased the caption as:
"The ensemble mean of daily streamflow from simulations (red), USGS observations (blue), and the daily 95% uncertainty range (grey) from 50 simulations with NSE greater than 0.5. The periods of (a) calibration was from Oct. 2012 to Sep. 2015 and (b) validation from Oct. 2015 to Sep. 2017. All simulations include irrigation, lawn fertilization, and septic processes"

l. 371 Caption figure 4 is instead of was

Thanks for the comment. We changed the "was" to "is".

l. 339: replace "turned on" with include

Thanks for the comment. We changed "turned on" to "include".

l. 465 Table 4 caption: I suggest for clarity to replace "others" with "each scenario"

Thanks for the comment. We rephrased as:
"Absolute and relative changes between scenario none and other scenarios are included below denitrification rates."

l. 482 typo: transform

Thanks for the comment. We changed to "transformation" here.

l. 494: it would be interesting to read here for how many years you suggest to spin up the model instead

Thanks for the comment. We mentioned that, in a previous RHESSys study (Lin et al., 2015), the spin up period was set to 500 years. However, the model was for a fully forested watershed where N input rate is much lower than our suburban watershed. With greater N input and a shorter residence time of N, we think spinning up model for 30 years resulting in stable C:N ratio, though longer spin up period could have more accurate results.

Line 498:

"A longer spin up period (i.e., 500 years) was used in Lin et al. (2015) for a fully forested watershed. In our suburban watershed with larger inputs and shorter residence time of N, the spin up period could be shorter than a fully forested watershed, as evidenced by asymptotic C:N ratio after 30 years."

l. 510: I like that you call it N retention hot spots here in the title, I suggest to consider to use this term throughout the text to differentiate more from N input hot spots which are described, too (Sec 2.4). This can increase clarity for the reader.

Thanks for the suggestion. We thoroughly checked the manuscript, and denoted "N input" and "N retention" in front of "hot spots" to further differentiate different types of hot spots. Here are some changes we made:

Line 33:

"The highest predicted denitrification rates, or N retention hot spots, were downslope of lawn and septic locations in a constructed wetland […]"

Line 266:

"[…] areas receive additional water and N input from septic effluents and may become N input hot spots of NO_3^- in the watershed."

l. 545 missing "of"

Thanks for the comment. We added "of" in the sentence.

l. 573-576 this last sentence is very long and redundant in some parts of it content

Thanks for the comment. We revised the long sentence in line 579 as:

"Our results showed the spatial pattern of N retention and identified spontaneously existing (accidental) retention zones that accumulate both water and N loads from upstream. By effective siting of BMPs based on our results for developed watersheds, both naturally occurring and built features could become N retention hot spots and provide ecosystem services to improve water quality in the future."

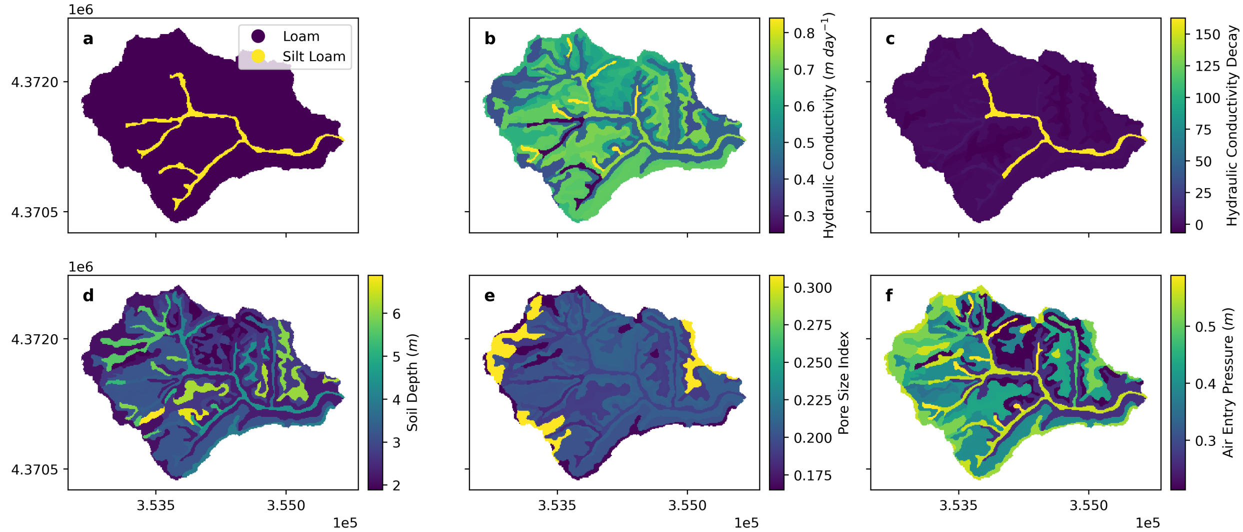
l. 578-581 Please divide this first very long sentence into several I suggest to improve the conclusion. There are some redundant points and very long sentences covering 3 to 4 rows.

Thanks for the suggestion. We rephrased this long sentence in line 584 as:

"Our analysis provides important insights into how different sources of N input interact with ecohydrological processes to control N export from suburban and exurban watersheds relying on local groundwater for domestic use and septic systems for wastewater release. With single-family houses dominant in these watersheds, the input of lawn fertilization and irrigation water, and septic effluent volume and N load are concentrated in limited areas at much higher per unit area rates."

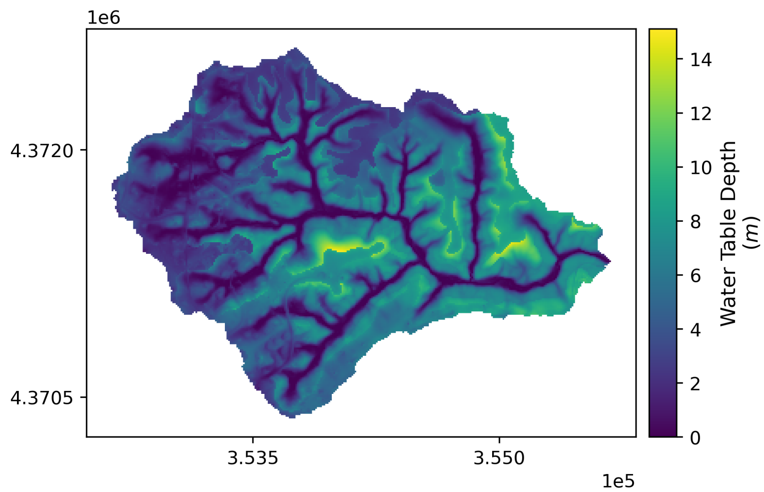
Figure A2: what values does the legend in panel a for soil texture represent?

Thanks for the suggestion. We labeled the soil texture in the figure, instead of using the index for soil type, to avoid the confusion.



l. 620: Figure A4 why (meter) and not (m)?

Thanks for the suggestion. We have changed the meter to "m" in the figure.



l. 645: The contribution of one author is not described. Please make sure all authors contributions are described in concordance with the CRediT contributor roles taxonomy

> Thanks for the comment. We included the contribution of Dr. Laurence Lin here for helping develop and calibrate the RHESSys model in line 661.
>
> "LL provided technical assist on developing and calibrating the RHESSys model, and PMG, AKS, JMD, and AJG provided water chemistry and biogeochemical data. All authors reviewed and edited the paper."

Report #2 (Referee #3)

Major comment: My main concern is the lack of rigor in model calibration.

First, though the model was evaluated on water and nitrogen dynamics, the model was only learned against the streamflow observations with limited parameters. Since N is the focus of the study, the parameters used in the ferritization and the septic system (i.e., Eqs (1)-(4)) should also be estimated based on the N observations. Otherwise, while focusing on transport and transformation, only the transport part of the model was well represented, which logically does not follow through.

Second, the flow calibration might be problematic. As major components of the water cycle, the ET and irrigation processes are not calibrated. Both potentially contributed to the underestimation of the low flow period in Figure 3. Although the authors admit this caveat in the discussion, it is suggested to tune the parameters of the two processes and discuss how the underestimation might affect the conclusion if unsuccessful.

Third, there is a lack of details in the current model calibration in Section 2.2, including
- I believe the authors generated an ensemble of model simulations and picked up the best run. If so, how many runs were performed? How was the parameter ensemble generated?
- Did the authors split the observed time period into training and test periods? It is recommended that this is done to limit the adverse impact of temporal extrapolation.
- There are two gages in the catchment (Lines 145-147). Have both gages been used to estimate model parameters? Or only the USGS one?

> ***Response:***
>
> Dear reviewer, we appreciate your time and suggestions to improve our model. Our manuscript focused on the transformation of NO_3^- through denitrification and evaluated the changes of terrestrial denitrification at many locations, including hot spots at riparian areas and wetlands (Figure 5 and Table 4). For your reference, all our responses are in blue color. We also quote and shade texts in the manuscript, and use red color to highlight new changes we made. We also carefully revised the entire manuscript to ensure that there is no grammar mistake found.
>
> For **calibration**, we substantially revised our calibration approach in our revised manuscript, compared to the initial submission which we believe was the version accidentally assigned to your review. Here are some main improvements we had in the revision:
>
> Firstly, in the revised manuscript, we included 50 behavioral simulations, which yield 1) NSE of streamflow from water year 2013 to 2015 greater than 0.5 and 2) gw2 lower than 0.5, to quantify the uncertainty of ecohydrological responses of our model (see the

uncertainty ranges in Figure 3 and 4). This improvement was based on the constructive comments from reviewers in the first round.

Secondly, there are no ET and irrigation observations that allow us to perform calibration for related parameters in this watershed. We would like to further calibrate the ET and irrigation in the future once high-resolution data are available. Otherwise, we don't want to overfit our model on streamflow data, which reduced the model's degree of freedom and increased bias potentially.

Lastly, we wanted to highlight that one significant aspect of our study is to restrict our calibration on soil and subsurface hydraulic parameters only. This allowed us to evaluate how well our model could be generalized to assess N dynamics for most ungagged suburban watersheds where only discharge, but no water chemistry (expensive to obtain) observations, are existing. We discussed this in Introduction and Discussion:

In Introduction (Line 103):
"The framework should be capable of extension 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, while restricting calibration to hydrologic parameters. Calibrating nutrient dynamics may not allow generalization to watersheds without chemistry records or extrapolation to conditions in which water quality BMPs are implemented."

Other minor revisions:
Line 34: best BMPs --> BMPs

Thanks for the comment. We deleted the "best" in our revised manuscript.

Figure 1: Please show the locations of POBR, the two gages, and the riparian area.

Thanks for the comment. The pond branch (POBR) is outlined in grey, and two gages are shown as red stars in Figure 1. Including the riparian areas (defined as height above nearest drainage, HAND, lower than 1.5 m) would make Figure 1 quite busy. The riparian areas are basically identical to dark blue (saturated) areas in Figure A4.

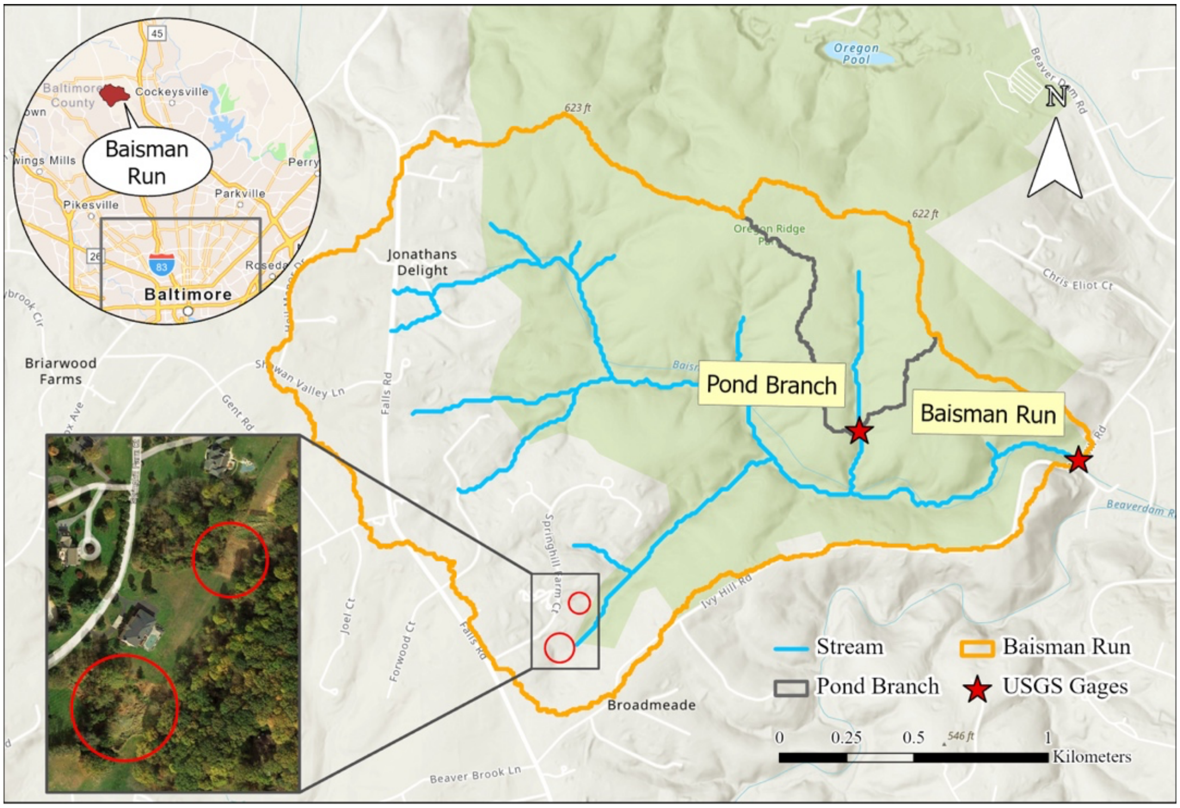


Figure 1. Study watershed Baisman Run (BARN) in suburban Baltimore County, Maryland (from ESRI). The inset highlights two N retention "hot spots": A sediment accumulation zone (upper circle) receiving drainage from roads and a constructed wetland (lower circle). Pond Branch (POBR), a fully forested subcatchment is outlined in grey.

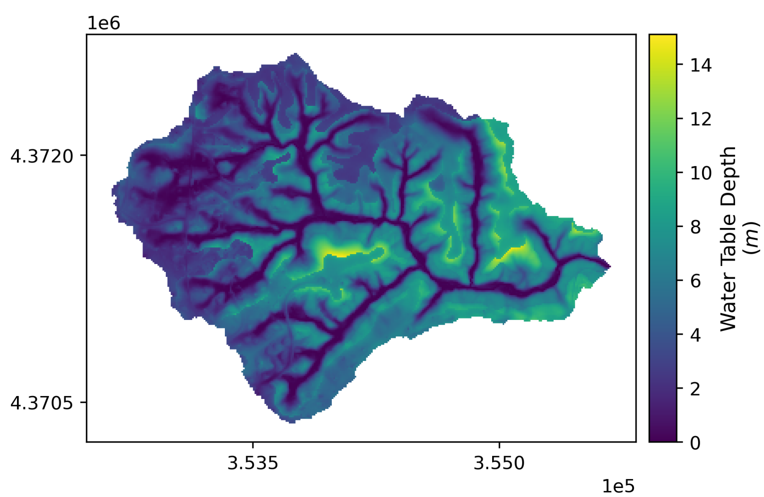


Figure A4. Spatial pattern of ensemble mean water table depth (m) of Baisman Run during the entire study period (water year 2013 to 2017) from the 50 behavioral simulations under scenario *none*. Map in projection NAD83 UTM 18N (EPSG: 26918).

Lines 152-153: Please present the spin-up storage of soil C and N.

Thanks for the comment. The ratio of soil C to N at the entire watershed was stabilized to 8.5 after the spin-up. We added this information at line 179:

"Inspection of the spin-up storage of soil C and N showed they were asymptotic with stable C:N ratios, with a mean of 8.5 in the entire watershed."

Section 2: Please provide a conceptual diagram and the key mathematical representation of nitrogen cycling (e.g., denitrification) used by RHESSys and link it to the proposed fertilization and septic system.

Thanks for the suggestion. The details of nitrogen cycling of RHESSys has been documented in Figure 2 of Lin et al. (2015). We mentioned in the manuscript to refer readers to check this diagram:

Line 254 & 284:

"Once NO_3^- is released to soil, N cycling is simulated following the procedure detailed in Lin et al. (2005)."

"Once NO_3^- is added to surface detention storage, N cycling is simulated following the procedure detailed in Lin et al. (2005)."

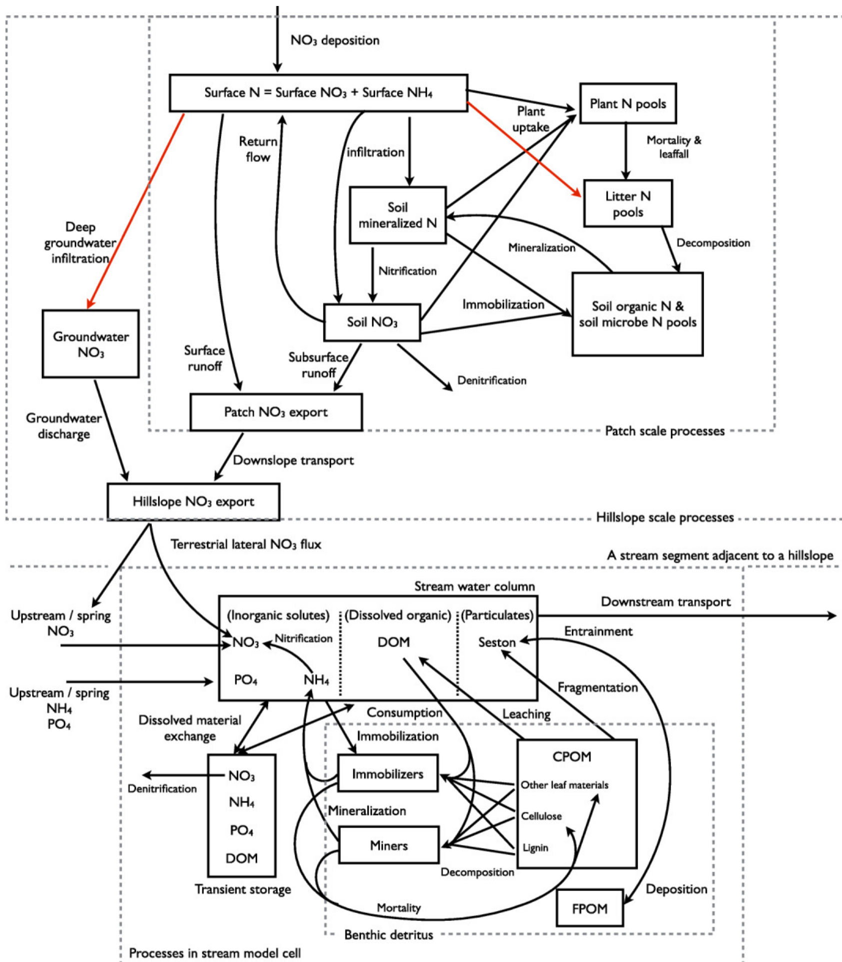


Figure 2 from Lin et al. (2005) showing detailed N cycling in RHESSys

Table 1: What are s,sv, svalt, and gw abbreviated for? Please illustrate.

Thanks for the comment. We included the physical meanings of these abbreviations in Table 1 as below:

Table 1. RHESSys parameters being calibrated and their physical representations (Tague and Band, 2004). Calibrated results shown as ranges of multipliers to original soil properties (Fig. A2 & A3) and groundwater component generating behavioral 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		-	0.51–1.98

		K_{sat0_v}	Vertical saturated hydraulic conductivity at the soil surface	USDA SSURGO, 2019	m day^{-1}	0.52–1.98
Soil properties	sval t	b	Pore size index	USDA SSURGO, 2019	-	0.51–1.98
		φ_{ae}	Air entry pressure		pounds inch^{-2}	0.5–1.05
Groundwate r 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

Lines 190-191: The authors state that "Both surveys were conducted during significant drought conditions (2002 and 2008) when lawncare was reduced due to groundwater supply concerns". Should irrigation be increased to account for drought conditions?

Thanks for the comment. Since there is no water supply sewer serving the watershed, groundwater is the only source for domestic water use. During severe drought, Maryland government may enforce water use restrictions (https://mde.maryland.gov/programs/water/droughtinformation/pages/restrictions.aspx), and homeowners may use water for prioritized purposes (e.g., sanitary and drinking) rather than irrigating lawns.

Eq.(2): The authors set a limit to the available irrigation water based on GW storage in the below cell. Should it be represented by a pipe-based irrigation system that extracts water from the entire GW pool instead of the below GW? That would change the result of water distribution, though.

Thanks for the comment. In our study watershed, water extraction was implemented at individual households. No center or piped irrigation system (or external water supply) exists in this suburban watershed. Also, RHESSys assumes a uniform groundwater storage in each hillslope (total of 16 in BARN, Figure A7), and irrigation water is extracted from there if one identified well is located within it. Lastly, as mentioned above, this irrigation module requires further data and study to be improved to better capture the practices at household level.

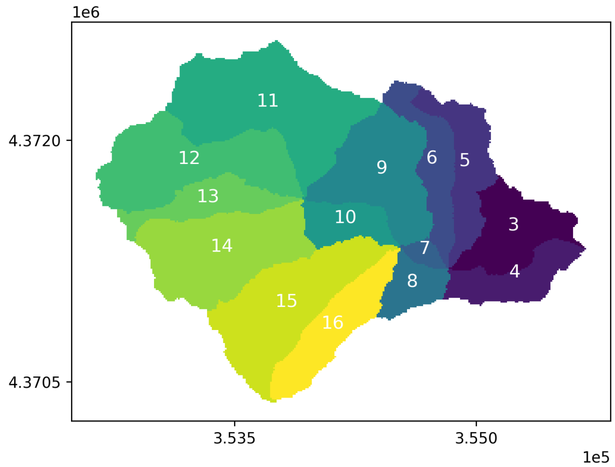


Figure A7. Hillslope indices of Baisman Run. Map in projection NAD83 UTM 18N (EPSG: 26918).

Eqs.(1) and (4): It is unclear how the authors set the parameters LF and IRmax. See my main comment above.

Thanks for the comment. Since Leaching Fraction (LF) could be confusing to readers, we changed it to release fraction (RF) in equation 1. Our model estimates RF by assuming 1) an exponential decay of fertilizer (similar to commonly found slow-release fertilizer) and 2) 10% of input fertilizer would remain after one fertilization interval (FI). This assumption could be overwritten if users have observations of fertilization practices.

Line 247:

"Assuming all lawn fertilization is done with the slow-release fertilizer designed to remain 10% after one fertilization interval, the daily release fraction (RF) is determined by the fertilization interval (FI), following Eq. (1):

$$RF = -\frac{log\ 0.1}{FI}, \quad (1)$$

In our case study, our 60-day fertilization interval results in 3.8% of nutrients in the fertilization pool to decline exponentially and transported to other pools per day and then stored, consumed by vegetation, immobilized, denitrified or further transported to groundwater and downslope. User-defined fertilization time series could overwrite this setting of lawn fertilization if observations are available."

The maximum daily **irrigation** was based on EPA's recommendation of about 1 inch per week (or ~4 mm per day) of irrigation.

We added the reference in our manuscript at line 301 as:

"In the current model, we defined the maximum irrigation rate (IR_{max}) in BARN as 4 mm day^{-1}, which was converted based on the EPA's recommendation (U.S. EPA, 2024) of one inch per week for lawns. This rate can be modified based on the local practices or for sensitivity analysis."

Lines 282-283 and Figure 3: The underestimation of low flow is during irrigation season. So, it could be attributed to either irrigation or ET, or both. A rigorous calibration of the two processes is needed. See my main comment above.

Thanks for the comment. As we mentioned above, there are no ET or irrigation dataset available in the watershed to calibrate our model. Calibrating additional parameters could introduce overfitting and reduce the degree of freedom (ability of generalization) of our model. We would like to improve the model's performance on ET, irrigation, and streamflow in the future once we have high-quality data in our study watershed.

Lines 291-292: How was the "mean streamflow-weighted long-term observed concentration" calculated? Please provide the equation.

Thanks for the comment. In our revised manuscript, we no longer use the streamflow-weighted concentration, but the absolute concentration (Figure 4).

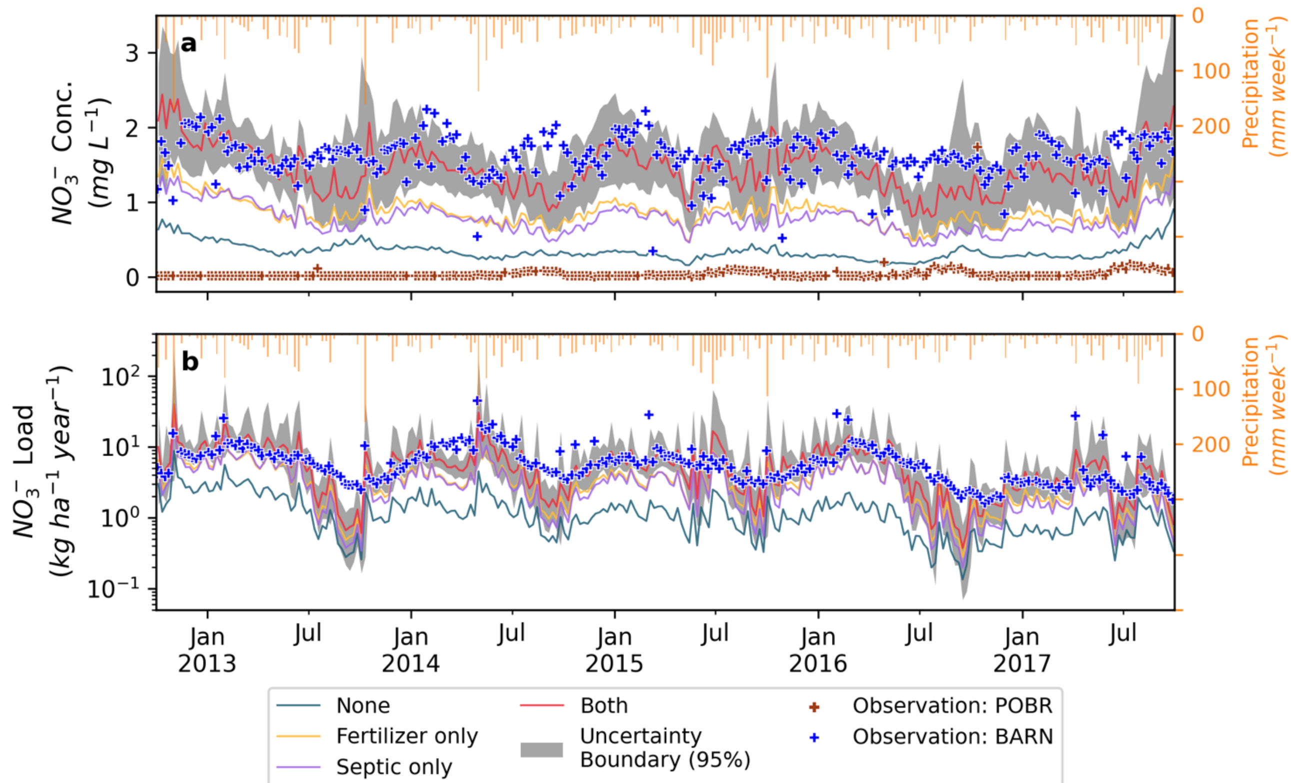


Figure 4. Ensemble weekly mean of (a) NO_3^- concentration and (b) load at the outlet of Baisman Run over the entire study period (water year 2013 to 2017). The 95% uncertainty boundary for scenario *both* is shown in grey.

Figure 5: Both nitrate load and concentration were underestimated for all cases. Again, the transformation is not calibrated.

Thanks for the comment. We apologize for showing the incorrect plot in our initial submission. Please refer to the new Figure 4 (shown above) to see how our model performed on simulating NO_3^- dynamics.

Section 3.3.2: Please provide the conceptual diagram of denitrification and its linkage to fertilization and the septic system.

Thanks for the comment. Please refer to the conceptual diagram, shown in our respond to the comment for "Section 2".

Figure 6: The difference (right row) is difficult to tell. Please either reduce the color bar range or use a different colormap to better illustrate the difference.

Thanks for the comment. We revised the figure as below:

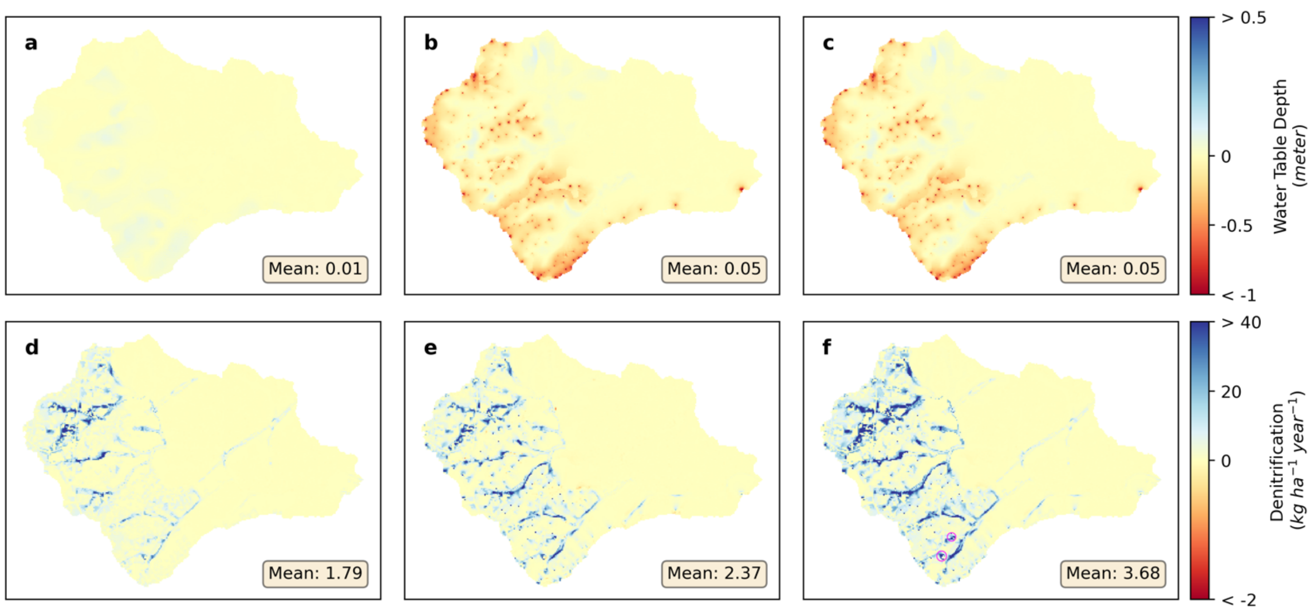


Figure 5. Ensemble mean differences of water table depth (top panel) and denitrification (lower panel) between scenario *none* and scenario *fertilizer only* (a & d), *septic only* (b & e), and *both* (c & f). The two hot spots of denitrification (i.e., wetlands in Fig. 1) were circled in (f).

Line 369: "… were in spring when fertilizer is applied to lawns and soil moisture is generally higher" --> "… were in spring when fertilizer was applied to lawns and soil moisture was generally higher"

Thanks for the suggestion. We used present tense here to show that the time of fertilization is generally in spring when soil is commonly wet. Using past tense could introduce the confusion that fertilization now is no longer initially applied in spring.

Line 370: "… had significant increases in denitrification in winter when the watershed receives" --> "… had significant increased in denitrification in winter when the watershed received"

Thanks for the suggestion. We used present tense here to show that riparian areas generally receive more NO_3^- during winter, due to stronger subsurface flow (no ET) and low N retention rate (low denitrification).

Sections 4.1 and 4.2: The authors discussed the underestimation of both low flow and nitrate in multiple places and attributed it to uncertainty in ET/nitrogen processes without calibrating them. Please see my main comment.

Thanks for the comment. As mentioned above, we wanted to keep the calibration simple (only on limited parameters) considering ET and water chemistry data are much less available than discharge for most gaged watersheds. Calibrating additional parameters could introduce overfitting and reduce the degree of freedom (ability of generalization) of our model.

In addition, we also wanted to evaluate whether our model can reasonably capture N dynamics assuming we only have discharge data that could be used for calibration. Our results of NO_3^- concentration and denitrification showed the reliable capacity of our model to be applied to vast watersheds without water chemistry data.

Report #3 (Referee #4)

Overall Comments:

Response:

We appreciate the referee's thorough comments to our revised manuscript. It was your constructive suggestions that help to improve our manuscript significantly. We have carefully addressed all comments below. For your reference, all our responses are in blue color. We also quote and shade texts in the manuscript, and use red color to highlight new changes we made. We also carefully revised the entire manuscript to ensure that there is no grammar mistake found.

1) Please clearly state, if accurate, that the simultaneous consideration of fertilization, septic systems, and irrigation schemes in the RHESSys model is novel to this study. Emphasize this point in both the abstract and introduction.

Thanks for the suggestion. Our augmented RHESSys model developed new modules of fertilization, septic, and irrigation processes at household/parcel level. Other RHESSys versions allow users to provide time series of irrigation and fertilization, but not at household level. Our model allows to simulate these processes more realistically for suburban watersheds like Baisman Run. We emphasized these in both Abstract and Introduction:

In Abstract (line 27):

"We augmented a distributed ecohydrological model, RHESSys, with estimates of the spatial distribution of these loads at household parcel level to develop a predictive understanding of the factors generating upland and riparian nitrogen cycling, transport and stream NO_3^- concentrations."

In Introduction (line 112):

"In this study, we augmented RHESSys to include household-level transfer of groundwater for lawn irrigation […]. […]. In summary, by adding modules of household-level lawn irrigation, fertilization, and septic releases (see Sect. 2.3) that are commonly found in suburban areas, RHESSys is designed […]."

2) Regarding the irrigation scheme, I am curious about its accuracy. Typically, farmers do not irrigate based solely on PET/ET conditions but rather follow conventional practices. Additionally, some areas may require more or less irrigation for various reasons. It would be beneficial if the authors could demonstrate the feasibility of this irrigation scheme by comparing it with satellite-based irrigation detection methods (though this is not mandatory) or by citing reliable sources. This is especially important given that the irrigation scheme is applied throughout the study period.

Thanks for the suggestion. Our current method (using ET/PET) is approximating the irrigation demand while considering the spatial distribution of lawns and households in watersheds. We discussed that the method cannot consider the actual heterogeneity of irrigation practices of people, including irrigation amounts and timing:

"For irrigation, our model applies irrigation close to its maximum (4 mm day^{-1}) when water stress is high, but residents may not irrigate their lawns at these rates during drought to conserve groundwater, and may continue to irrigate lawns during wet periods with automated sprinkler systems."

Though we would like to better capture the irrigation signals at household level, there is no dataset available for irrigation at the resolution (< 100 m) we needed to account for the heterogeneity of irrigation practices in our watershed (just 3.8 km^2). In other words, we need more survey and high-resolution observations in the future to account for these household-level heterogeneity of irrigation. We added a sentence in the Discussion in line 543 to highlight this limitation:

"Surveys and high-resolution satellite observations could help to improve our irrigation module and accurately estimate the timing and quantity of irrigation practices in suburban watersheds."

Specific Comments:

Line 33: Remove "in the range of measured values" and include the exact numbers.

Thanks for the suggestion. The range of measured denitrification rates from Suchy et al. (2023) and Raciti et al. (2011) varies a lot. We would need more information to help readers understand these, and this would make our abstract too long. We think it would be better for reader to check the Section 3.3.2 for the details of denitrification.

Line 50: Did you consider spontaneously developed "hot spots" in the model in this study?

Thanks for the comment. We discussed the denitrification rates in the constructed and accidental wetlands (Fig. 1) were highest within the watershed at line 441:

"The annual denitrification rates in the sedimentation accumulation zone (upper red circle in Fig. 5) showed a significant increase after activating fertilization and septic processes, from 76.9 kg N ha^{-1} $year^{-1}$ before to 95.6 (+18.7, 24.3%) kg N ha^{-1} $year^{-1}$ after activation. Similarly, denitrification rates in the constructed wetland (lower red circle in Fig. 1) increased from 81.5 kg N ha^{-1} $year^{-1}$ before to 102.7 (+21.2, 26%) kg N ha^{-1} $year^{-1}$ after activation."

And we further discussed these locations as hot spots in the watershed at line 529:
"Specifically, these two wetlands covering only 0.09% of the watershed contributed to 0.39% of the total denitrification during the study period."

Line 108: Define "RHESSys" earlier in the manuscript, as it appears before Line 108.

Thanks for the comment. We moved the full name of RHESSys to line 89 as:

"Fully distributed hydrology models, such as […] RHESSys (Regional Hydro-Ecological Simulator System, Tague & Band, 2004) could explicitly […]"

Line 134: BARN has already been defined; there is no need to define it again.

Thanks for the suggestion. We revised the sentence in line 134 as:

"Our study watershed (Fig. 1), BARN, is in Baltimore County, MD, […]"

Line 145: Is the temperature range accurate? Is there no snowfall?

Thanks for the comment. Baltimore has a moderate and wet climate. The mean annual maximum and minimum temperatures are as reported. Snow does occur in Baltimore in the wintertime.

Line 155: Are there other facilities, such as tennis courts or soccer fields, in this area that could alter the land properties?

Thanks for the comment. Except for forest, sparse single-family houses are the dominant land use in BARN. No large facilities, such as schools or sport fields, are found in the watershed by inspecting Google Earth satellite imageries.

Line 179: Please specify the ratio mentioned.

Thanks for the comment. The ratio of soil C to N at the entire watershed was stabilized to 8.5 after the spin-up. We added this information at line 180:

"Inspection of the spin-up storage of soil C and N showed they were asymptotic with stable C:N ratios, with a mean of 8.5 in the entire watershed."

Line 205: Define SSURGO before using the term.

Thanks for the comment. We added the full name of SSURGO here as:

"[…] initial estimates (Fig. A2) from the Soil Survey Geographic Database (SSURGO, USDA, 2019)"

Line 218: Please specify these limits in the supplementary information or provide references.

Thanks for the comment. We added the reference, Smith et al. (2022), here in line 206.

"From four thousands of parameter set realizations randomly chosen within specified limits described in Smith et al., 2022, behavioral sets are chosen as yielding Nash-Sutcliffe efficiency […]"

Line 255: Provide a reference to confirm that all households in BARN use septic systems. Do they all use the same system? How do you ensure that the septic systems were functioning properly during the study period?

Thanks for the comment. BARN is not within the Baltimore's Urban Rural Demarcation Line (Zoning regulations of Baltimore, https://perma.cc/LCC2-8BZJ), outside which no centralized sewer systems, including water supply and sanitary sewage, exist. Therefore, all households use their own septic systems for wastewater drainage.

As each household heavily relies on the septic system for daily life, we believe homeowners would quickly fix issues if septic systems were not functioning. As the result, our study did not consider these incidents, assuming the impacts of short infunctioning are negligible.

Line 297: Explain why 4 mm/day was chosen as the maximum irrigation rate in BARN. Any references?

Thanks for the comment. The maximum daily irrigation was based on EPA's recommendation of about 1 inch per week (or ~4 mm per day) of irrigation. We added the reference in our manuscript at line 301 as:

"In the current model, we defined the maximum irrigation rate (IR_{max}) in BARN as 4 mm day^{-1}, which was converted based on the EPA's recommendation (U.S. EPA, 2024) of one inch per week for lawns. This rate can be modified based on the local practices or for sensitivity analysis."

Line 341: ?

Thanks for the comment. The $\pm$ symbol was not shown correctly. We have fixed these syntax issues.

Line 350: Provide an explanation for the underestimation of streamflow from July to October 2016. Additionally, please include the time series of WSF values in the supplementary information.

Thanks for the comment. For the streamflow underestimation, we discussed at line 467 as:

"We also noted that our model tended to underestimate the lowest streamflows during the growing season, which was also found in another suburban watershed, Dead Run, in Baltimore by Miles (2014). Several potential reasons could cause this discrepancy: 1) Higher transpiration estimates caused by uncertainties in vegetation ecophysiological parameters in RHESSys controlling vegetation water use or phenology; 2) Underestimation of groundwater recharge and release to streams during the growing season; and 3) A lack of household modulation of groundwater use during dry periods. During our prior surveys (Law et al., 2004; Fraser et al., 2013) residents stated they had reduced their water use during droughts. While the model underestimation was negligible, additional empirical data about water flux, groundwater processes, and household water management would enhance model prediction accuracy of hydrological processes, especially during the growing season."

We included the watershed-scale water stress factor at each patch (Eq. 3) as Figure A6:

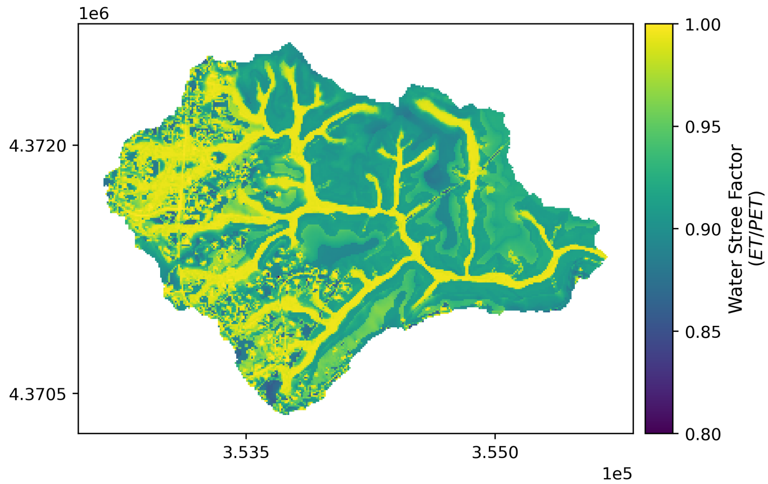


Figure A6. Spatial pattern of ensemble mean water stress factor (ET/PET) of Baisman Run during the entire study period (water year 2013 to 2017) from the 50 behavioral simulations under scenario *none*. Map in projection NAD83 UTM 18N (EPSG: 26918).

Figure 5: Consider adding arrows to highlight the circles, as they are difficult to see.

Thanks for the comment. We added an inset to highlight the two wetlands (circles).

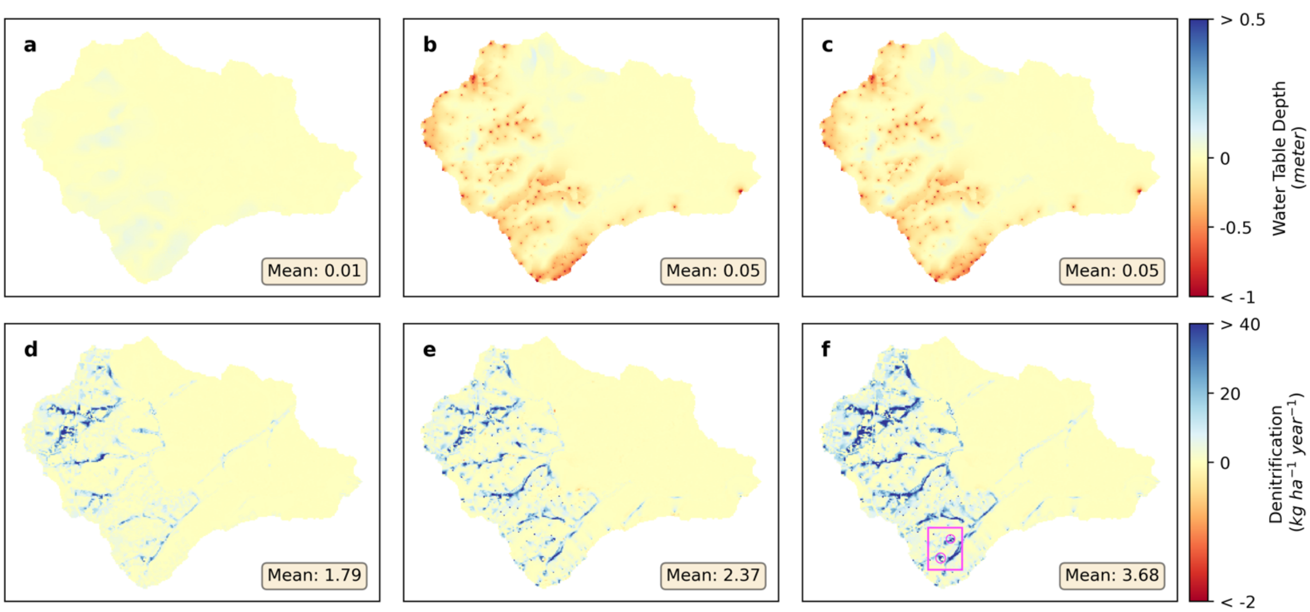


Figure 5. Ensemble mean differences of water table depth (top panel) and denitrification (lower panel) between scenario *none* and scenario *fertilizer only* (a & d), *septic only* (b & e), and *both* (c & f). The inset highlights two hot spots of denitrification (i.e., wetlands in Fig. 1) were circled in (f).