Report #1

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

In the revised version of the manuscript, the authors clarified the methods used in Section 3 and added a discussion about other sources of uncertainty beyond climate and land use in Section 5.2. The analyses performed and the results remain unchanged. I think that the manuscript has significantly improved in terms of clarity. However, I still have a number of comments, in particular regarding the nitrogen and climatic input data and the derivation of the value of 10% for the uncertainty in the reduction map. In addition, although the authors did not assess the uncertainty coming from the model parameters, the manuscript highlights other sources of uncertainty (bug in Mike She code or model error in validation) that are of the same order of magnitude as the uncertainty due to climatic/land use conditions (10%). I think that this should be better discussed.

Importantly, I would like to make a clarification with respect to the authors' reply to my second comment that 'A global search engine (Shuffled Evolution Complex) was used. Thus, limiting the risk of equifinality'. I would like to highlight that this statement is not correct. Equifinality results from a lack of information/data to constrain the parameters of a given model, and does not depend on the method used to estimate the parameters. Optimization method such as the Shuffled Evolution Complex algorithm may not reveal equifinality because they identify only one parameter set that 'best' fit the data, while there is likely to be other parameter sets with similar performance that have very different parameter values. I refer e.g. to Beven et al. (2006) for a discussion on the issue of equifinality.

Thank you for this comment. Yes, you are of course completely right. We meant to say that the risk of landing in a local minimum was reduced by the global search engine. This is not equifinality, this was entirely a mistake from our side.

We have made the discussion on the uncertainty clearer, e.g., by introducing a new section (5.3), where all reflections on uncertainties and limitations are gathered from the other parts of the discussion. We have also added and modified sections on the above issue on split-sample and equifinality in L645-655 and L679-692:

L645-655: When using a hydrological model for simulating impacts of changes in catchment conditions compared to those existing in the calibration period, split sample validation tests are not sufficient to document a model's capability to simulate hydrological changes. Experience shows that models that are used for making predictions beyond the conditions for which they are calibrated, such as land use or climate change in the present study, due to equifinality often suffer from model structural uncertainties (Refsgaard et al., 2012). In such situations the more comprehensive and data demanding differential split sample tests are recommended (KlemeŠ, 1986; Refsgaard et al., 2014). Due to lack of data such tests were beyond the scope of the present study. Instead, the model structural uncertainty for the present case was assessed using a multi-model approach with two additional hydrological models (Karlsson et al., 2016) suggesting that the signal coming from climate change was dominating over model structural uncertainty as far as hydrological change is concerned. Therefore, we argue that the inevitable uncertainties arising from model use beyond calibration conditions most likely are not so large that they affect our conclusions.

L679- 692: During calibration and validation of the model, a decrease in model performance was registered in the water balance for the validation period. This could be caused by non-optimal parameter estimates, and there is always a risk of equifinality during model calibration. In this case, the risk was minimized using *an extensive dataset of both discharge, hydraulic head, redox depth and nitrate flux during calibration of the different model steps. However, a multi model set as used in this study may still be prone to the risk of equifinality. Parameter estimation are here done in a stepwise fashion for each of the models, and the catchment scale calibration of DAISY along with the particle tracking approach limits the evaluation of performance of the nitrate component to mean catchment values. The dynamic of the nitrate system is thus impossible to verify. To account for this a full solute transport solution would be necessary but was unfortunately not possible in the framework of this study; but would be a relevant next step in investigating uncertainties and improve model verification.*

During simulations it was found that some particles were not correctly released due to model error, and we therefore were forced to make the assumption that the particles would be distributed similar to non-trapped particles. However, the validity of this assumption is associated with considerably uncertainty. The correction led to mean changes in reduction of 2 %, but the resulting impact on the true reduction map is unknown, as we do not know the actual travel path of the trapped particles.

KlemeŠ, V., 1986. Operational testing of hydrological simulation models. Hydrological Sciences Journal, 31(1): 13-24. DOI:10.1080/02626668609491024

Refsgaard, J.C. et al., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. Advances in Water Resources, 36: 36-50. DOI:https://doi.org/10.1016/j.advwatres.2011.04.006

Refsgaard, J.C. et al., 2014. A framework for testing the ability of models to project climate change and its impacts. Climatic Change, 122(1): 271-282. DOI:10.1007/s10584-013-0990-2

I report below detailed comments.

p9 L203: Could the authors briefly explain what the crop recommended nitrogen rate is and where the value come from (reference)?

Thank you, we have added the following statement: *The crop recommended nitrogen rate based on soil type and crop sequence from Danish Ministry on Agriculture (Plantedirektoratet, 2005) for the years 2004-2007 was used to setup the fertilization scheme.*

p9 L210: Could the authors add the source of the nitrogen atmospheric deposition dataset.

Thank you for this comment. The source is given at the end. We have tried to indicate this better by moving the reference: *For crop rotations including clover grass and peas nitrogen biological fixation is calculated using Høgh-Jensen et al. (2004) and nitrogen atmospheric deposition is included as input to the soil using standard DAISY settings (given in Hansen et al., 2012b) for dry and wet deposition.*

p12 L279: The calculation of potential evapotranspiration is still unclear. What is 'water vapour' and 'water pressure'? I guess that the authors mean vapour pressure/relative humidity and atmospheric pressure, which are the variables used in the Penman Monteith equation. In addition, more details are required on the radiation term. Net radiation is necessary to compute Penman Monteith equation, while typically climate models provide the downward radiation terms only.

Thank you for pointing out this shortcoming. We have changed the sentence to: *The reference evapotranspiration is calculated using FAO Penman– Monteith formula adapted by Allen et al. (1998) based on the climate model outputs for minimum and maximum temperature, incoming long and short wave solar* *radiation, relative humidity and wind speed. Following the recommendations in Allen et al. (1998) and Seaby et al. (2013) variables needed for the Penman– Monteith formula e.g., net radiation (calculated from the net incoming short and long wave radiation), water vapour pressure, height-adjusted wind speed and atmospheric pressure, where calculated from these outputs.*

p18 Table 4: I guess that the mean and standard deviation refer to the spatial distribution of the nitrate distribution potential? Please clarify.

We are not totally sure what is meant, but we have tried to make the table text of what is shown clearer: *Table 4: Mean and standard deviation (in brackets) across the catchment for each of the nitrate reduction potential maps (proportion of nitrate N reduced).*

p21 L421: Figure 4 does not report results on reduction potential. Please refer to the appropriate figure/table for this.

True this should be Figure 3. This has been corrected.

p28 L573-579: What about nitrate reduction in soils?

Yes, that is correct, we have not mentioned that the soil type could of course also influence the results. We have added a sentence on this in L657: *The present study was carried out for a groundwater dominated catchment characterized by till deposits, confined aquifers, and relatively shallow redox interfaces and phreatic groundwater tables. Furthermore, this catchment has a relatively uniform soil type distribution, dominated by clayey soils.*

p28 L594: Where does the value of 10% comes from? It is not clear from the results presented in Section 4 (Figure 6).

We have added a line on how this value is found in L570*: Across all climate models, the average absolute effect on nitrate arrival is 10%, when using a fixed reduction map compared to a targeted reduction map.*

p29 L602-603: I think that this uncertainty of up to 10% should be also discussed with respect to other modelling uncertainties that are reported in the manuscript and that are of the same order of magnitude, i.e. the uncertainty due to the bug in the Mike-SHE code (that is up to 9% p12 L267), and the model error in the validation period (that is up to 10% for Mike She p15 L328 and up to 23% for Daisy model p15 L333)

Thank you for this comment. Please see the reply for the main comment above.

p29 L612: There is something wrong is this sentence. Please clarify.

This sentence has been deleted and the section moved to L666.

P30 L628 'Thus, the uncertainty of the reduction maps is dependent on both model setup and assumptions': to which model setup and assumptions do the authors refer? The uncertainty to model setup and assumptions is not analysed in the study, but only discussed as potential uncertainty that may arise. This second conclusion should be amended to reflect this.

Thank you for this comment, we have changed the sentence according to the suggestion: *The magnitude of the changes in the reduction map found here may, however, be influenced by both model setup (e.g., drainage), model errors (e.g., particle flow paths) and assumptions (e.g., fixed redox interface). Furthermore, the span of the chosen land use and climate change scenarios analysed and the flow regime in the study catchment may also influence results.*

Minor edits

p3 L57 ‘have been applied’: what is the subject of ‘have’? Is it ‘a nitrate reduction map’? In this case it should be replaced by ‘has’.

Thank you, this has been corrected

p4 L93: replace ‘constitute’ by ‘constitutes’

Thank you, this has been corrected

p4 L94 ‘fellow’: Do the authors mean ‘fallow’?

Yes, this has been corrected

p4 L97: replace ‘as a results’ by ‘as a result’

This has been corrected

p5 L112: replace ‘in 100 meter’ by ‘at 100 meter’

This has been corrected

p6 L131: ‘inverse’ should be replace by ‘inversely’, but I would actually remove this term.

Thank you, this has been removed

p7 L147: replace ‘need’ by ‘needed’.

This has been corrected

p10 L224‘accumulated’: do the authors mean ‘cumulative’?

Yes, this has been corrected

p11 L239: replace ‘representable’ by ‘representative’.

This has been corrected

p11 L261: add ‘it’ after (‘the data from which’)

This has been corrected

p14-15 Table 2: I suggest to refer to ‘observational period’ and ‘reference period’ rather than ‘control period’

Thank you, this has been corrected

p16 Table 3 caption: replace ‘for the crop type’ by ‘for the crop types’

This has been corrected

p20 L404-405: the verb is missing is this sentence.

Verb added

p22 L437: remove the ‘s’ at the end of ‘results’ (in ‘these changes results’).

This has been corrected

‘p.p.’ is used multiple times in the manuscript: what does it stand for?

Thank you, yes we have not explained this. P.p. stand for percentage point, we have added the definition in Line 298

p25 L520: replace ‘figure 9’ by ‘figure 6’.

This has been corrected

p30 L624-625: replace ‘the main finding of the study was by ‘the main findings of the study were’

This has been corrected

References

Beven, K. (2006). A manifesto for the equifinality thesis. Journal of Hydrology, 320(1–2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Report #2

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

The authors have addressed the previous comments and have greatly improved their manuscript and its clarity. I have a few more, but minor comments that should be addressed before publication of this manuscript.

General (minor) comments:

- I appreciate the much clearer Method Sections that you greatly improved. Just, sometimes I feel it is not completely which methods were set up previously in other studies and simply reused for this manuscript or which ones you extended from previous studies. One example is the part on redox depth estimation paragraph L236-242. Maybe in some cases it would be helpful to start with the reference instead of at the end of the description. "Following …, we did…" or "We extended on the method … presented by ….". Please revise those sections to be clear. This is not easy as your work makes use of many previous approaches from the study region and in general, but I think it would further improve the readability and the connection to the references.

Thank you for this suggestion. We have tried to indicate this clearer in the method section in L147, L171, L201, L213, L232.

- Check consistency of spelling:

1.) NO3-N should all be NO3-N or "nitrate-N" ;

Thank you, this has been corrected.

2.) lower and upper case usage for each of the models MIKE SHE and DAISY;

Yes, this is inconsistent. We have corrected this to upper case for both models throughout the manuscript.

3.) When going through the manuscript I found mostly 2080-2099 for the future period, but in some instances 2088-2099, please revise to ensure these are correct.

Thank you, this has been corrected.

Detailed comments:

L 10: I still think this formulation is overly complicated. I suggest formulating the goal of your study in a positive way, something like investigating the "potential improvement of using transient nitrate reduction maps compared to …".

Thank you for this suggestion, we have added this in the text.

L 36: This sentence should be more precise in my opinion. Denitrification is one of the removal processes not just a synonym of the term, e.g. what about N uptake? Especially when you talk about processes in the unsaturated zone as you simulated with Daisy and in surface waters as you mention here, I think the terms should be introduced more carefully. It merits this attention for the aim of your study. I think the next paragraph would also fit to include "denitrification" term.

Thank you for this suggestion. We have amended the text to make the description more precise: *Nitrate is removed by a set of natural near-surface removal processes including plant uptake and soil retention,* *furthermore, the natural removal of nitrate in the groundwater and the surface water must also be considered, when assessing the impacts of nitrate leaching from agricultural areas on aquatic ecosystems. This removal, takes place via natural biogeochemical reduction processes often referred to as denitrification It can be expressed as a percentage removal and depending on the actual hydrobiogeochemical conditions, the denitrification may mainly occur in groundwater or in surface water systems such as lakes or wetlands (Huno et al., 2018; Quick et al., 2019).*

L 40: remove "(nitrate reduction)"

This has been corrected.

L 42 and 45: inconsistent spelling lower and upper case "quaternary"

This has been corrected.

L 55: An approach is singular, please revise

This has been corrected.

L 102: "the load is approximately …" What kind of load? Is this input as N surplus or the export in the stream? This should be more specific, and depending on your answer possibly also include how it is estimated, e.g. mean discharge times mean concentration?

Thank you, we have modified the sentence to: *The station, therefore, provides a long and near-complete data set for nutrient modelling as well as an extensive water discharge time series (Trolle et al., 2019). In 2005-2009, the average discharge amounts to 4.6m^3/s and the transport in the stream (load) is approximately 14 kg NO3-N/ha/year, calculated from measurements of mean concentration and mean water discharge.*

L 106: To start a sentence with a number is not a nice style. Plase revise

This has been corrected.

L125: Danish

This has been corrected.

L 155: I am curious: why did you select these periods for calibration and validation? Why do you use more years for validation (in total) than for calibration? From my knowledge this is rather uncommon, therefore I am curious about this decision.

The calibration period was chosen because this period has the best dataset, the first ten years (1990-2000) where used as warm-up. Subsequently, the validation period was then simply defined before and after the calibration period.

L231-235: Can these graphs be shown in the supplements? I think this could be good additional information. Also add reference to SI here.

Thank you for this suggestion, we have added the figure to the supplementary material and included a reference to the figure in the supplementary material in the text.

L 270: 16 models, but L 306 20 simulations. Is this correct?

Yes, this is correct. In the 20 simulations, there are also included the 5 runs with baseline land use. We have deleted this sentence as it is not necessary and causes confusion.

L 597: New inserts “N reduction” are not consistent with “nitrate reduction” in the rest of the text, recheck please

Thank you, this has been corrected.

L 615: “To account for this, a full solute” comma missing

Comma was added

Table 4: response to your response: “Table 4: Do you have an idea why the standard deviations of all models are that similar (Table 4, 0.36-0.39)? Can you comment on that, please?” – “This could be related to the fact that most changes in the reduction map are happening for values close or around the mean. Areas with 0% og 100% reduction are perhaps less likely to change reduction potential, as they are either very close to surface water (0% reduction) or located at areas with a deep groundwater level (long deep flow paths and limited drain flow) (100% reduction).” Maybe, that is what you meant. Actually, now with your response, I think that it is because all models cover the range of 0-100% with relatively high amount of cells in the extremes, stretching the standard deviation to similar values independent of the mean. I am not sure, the standard deviation is telling considering this perspective and the not normal distributed values. Figure 2b for example clearly shows that <10% and >90% prevail.

This sounds like a very valid reason for the similarity in the standard deviations. I think you are probably right, and therefore the standard deviations are maybe not so informative. We have chosen to leave them in, as it is difficult to test, if that is the reason in the actual case.

Report #3

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

The authors have made many changes in the revised version in an effort to address the comments from three reviewers. Unfortunately, they could not really address our concerns about equifinality, model calibration and uncertainty assessment, and perform a convincing evaluation of the model with observational data. The calibration of a parameter-rich model still relies on manual calibration of some parameters and with little use of observational data to evaluate the model. The research question "how changes in flowpaths and timing of nitrate release will influence denitrification in a context of global changes" is interesting, but the methods are relatively poor (although probably labour intensive).

While we agree that our model is parameter rich, we do not agree to the reviewer's statement that we have made "little use of observational data to evaluate the model". The reviewer statement about "concerns about equifinality, model calibration and uncertainty assessment" is very general without mentioning specific issues or problems. In the first round review, the reviewer expressed only two specific concerns (poor evaluation of interannual variability and need for differential tests related to land use and climate change) which we addressed in the discussion at that time. Altogether, we cannot follow the general statements expressed in the review, and we do not find the criticism of inadequate information on model calibration and evaluation fair, because:

- The modelling system used in the present study comprises three components: 1) The MIKE SHE hydrological model; 2) the DAISY root zone/nitrate model; and 3) the model describing the depth to the redox interface. Calibration and evaluation of the first two models (MIKE SHE and DAISY) was reported in a previously, well cited (92 ISI citations) journal paper (Karlsson et al., 2016). We do not think it is useful to repeat too much of the descriptions from the old paper, which we instead have referred to.
- For the hydrological model we used observational data from 415 groundwater wells and four discharge stations as calibration targets distributed across the 486 km^2 catchment. This is in our experience a relatively data-rich catchment suitable for the kind of analysis we have performed.
- It is correct that the hydrological model contains many parameters. However, through a rigorous sensitivity analysis including 43 parameters we ended up calibrating only the five most sensitive parameters. We used a global search parameter optimization algorithm reducing the risk of finding local optima. We believe that the hydrological model calibration and evaluation as reported in Karlsson et al. (2016) have been performed using state-of-the-art methodologies. Having said that we agree to the reviewer statement in the first review that it would have been useful with some kind of differential split-sample tests to evaluate the model's capability to simulate changes in land use and climate. As explained in the previous interactive discussions, such tests were not possible with the existing data. Instead the uncertainties in predictions of land use and climate change have been assessed using a multi-modelling approach in Karlsson et al. (2016). We agree that this issue has not been adequately addressed in the previous manuscript and we have therefore added a discussion on this concern and its possible implications for the study conclusions in our revised manuscript.
- The calibration of DAISY has been performed using the standard protocols for this type of model (Styczen et al., 2004) as described in Karlsson et al., 2016).

- The third model component, which was not included in Karlsson et al. (2016) is the model used to calculate the depth to the redox interface. As explained in the manuscript, this is a relatively simple model with only two parameters used for calibration against total N-load at the catchment outlet. To reduce the equifinality, the final values of the two parameters were fixed by comparing the calibrated redox interface to independent data on depth to redox interface collected from 226 wells. The information from these wells come from the national database that according to legislation stores information on all water supply wells in the country. According to our experience 226 wells with redox information within a 486 km^2 catchment is a quite high data density, and we cannot see how we possibly could have obtained better data or methods for evaluating the simulated depth to redox interface.

Based on the above we would argue that our modelling methodology basically is sound and suitable for the purpose of the study. We can agree to two of the specific concerns and have modified our discussion in the revised manuscript with respect to:

- We agree that some kind of differential split-sample test would have been useful to evaluate how good the model is in simulating effects of changes in land use and climate. But we argue that this aspect has been adequately evaluated by the multi-modelling study performed by Karlsson et al. (2016).
- We agree that our set of models inevitably includes some equifinality.