Reply to reviewer comments on hess-2018-375

January 18, 2019

Dear Dr. Christian Stamm,

herewith we submit the revised version of the manuscript hess-2018-375 'A comprehensive sensitivity and uncertainty analysis for discharge and nitrate-nitrogen loads involving multiple discrete model inputs under future changing conditions'. First we want to thank you for handling the manuscript and more importantly Francesca Pianosi, Björn Guse, and the two anonymous reviewers for their detailed reviews. We highly appreciate the constructive comments provided in the reviews that, without doubt, improved the quality of the manuscript. We tried to consider all comments in the revised version of the manuscript with the best of our knowledge. Below we added the all reviewer comments, our suggestions to revise the manuscript, and the references to the sections in the updated document, as well as a document indicating the differences between the revised and the previous version of the manuscript.

We hope the revision of the manuscript consider all comments made by the reviewers accordingly. If there are any further questions or any further issues from our side to handle, please contact me and we will try to clear them as soon as possible.

Sincerely,
Christoph Schürz

Reply to the reviewer comments RC1: 'Review of the manuscript by Schürz et al.' by Björn Guse

Summary

In this manuscript by Schürz et al., a detailed analysis of the impact of scenario simulations on hydrological variables is presented. In a sensitivity analysis, the sensitivities of five groups are separated. These groups are three types of scenarios (land use, point source and climate) and two model-specific groups (model set-up, model parameterisation). In the analysis, the impact of the input variables and of the uncertainty of the selection of scenarios or model characteristics is presented.

Overall, I like the manuscript. However, I see still potential for improvements to increase the understanding of the manuscript.

We would like to thank Björn Guse for his very constructive review and the valuable comments to improve the quality of the manuscript. We appreciate the positive feedback on the manuscript. In the following, we addressed each comment individually. The comments made by Björn Guse are printed in *serif, italic font*. Our replies to the comments are written in black, non serif font and our suggestions to revise the manuscript according to a comment are highlighted with the colors blue for insertions and red for deletions. The reference to the final modification in the revised manuscript is given below each reply to a comment in green.

Major comments

From my perspective, the readability of the manuscript can be increased by a clear separation in the two major points of the article. First, the impact of the input variables is analysed to show which input variables are more relevant for discharge and nitrate. Second, it is analysed how the selection of a scenario or model characteristic controls the target variables (uncertainty analysis). I think that the article would be easier to understand if these two aspects are clearly separated. This comment is mainly related to abstract, introduction and discussion. In contrast, these two aspects are already clearly separated in the conclusion.

We strived for a consistent structure in the manuscript by presenting the two separate blocks: 1) the sensitivity analysis of the model inputs and the model setup and parametrization; 2) and the uncertainty analysis together with the visual analysis.
With your comment in mind, we see the equivocality in the outline of the manuscript. In the current form the manuscript outline can be interpreted as if the performed sensitivity analysis was the actual focus of this work and the visual analysis of the uncertainties are a mere by-product (e.g. p.1L15ff in Abstract, p.5L7-16 in the Introduction).
We agree, that emphasizing the value of the visual analysis and treating it as an individual part of this work increases structure and consistency of the manuscript (e.g. being consistent to section 2.6 p.12L18ff where we already clearly separate the two goals of the performed analyses).

We suggest the make following changes in the updated version of the manuscript:
Exemplary suggested changes in the abstract, p.1L15ff:
The analysis of the 7000 generated model combinations of both case studies had two main goals; i) to identify the dominant controls on the simulation of discharge and $NO_3^- - N$ loads in the two case studies and ii) to assess how the considered inputs control the simulation of discharge and $NO_3^- - N$ loads. ~~In both case studies we employed global sensitivity analysis (GSA) t~~ To ~~identify~~ assess the impact of the input scenarios ~~scenario inputs~~, the model setup and the parametrization on the simulation of discharge and $NO_3^- - N$ loads we employed methods of global sensitivity analysis (GSA). The uncertainties in the simulation of discharge and $NO_3^- - N$ loads that resulted from the 7000 SWAT model combinations were evaluated visually. We present approaches for the visualization of the simulation uncertainties that proved to be a powerful diagnostic tool in this study to assess how the analyzed inputs affected the simulations. ~~We accompanied the GSA with a visual analysis of the simulation outputs and their associated uncertainties that resulted from the simulations of the 7000 SWAT model combinations. We present visualizations of the results of the GSA and the simulation uncertainty bands that proved to be powerful diagnostic tools in this study.~~

Following the suggested changes for the abstract we suggest to apply the same ideas for revising the introduction. Here we will focus on the sections p.3L21-29 and p.4L7-16 in particular.
Concerning the discussion we would prefer to keep the approach we currently follow in the manuscript, in which we discussed our findings and related them to other literature. Clearly separating the findings concerning the sensitivity analysis and the uncertainty analysis is difficult to facilitate for the larger part of the discussion. A separation might lead to a lot of repetition in the discussion.

We modified the sections p.1L15 - p.2L2 and p.4L15-26 accordingly

P.11, L: 1: Is it correct that you have identified 43 and 52 behavioural parameter sets out of 100.000 model simulations? If it is true than the number of behavioural parameter sets is rather low. How is than the impact on the sensitivity analysis meaning that most of the parameter sets are unbehavioural?

The decision whether a parameter set is considered to be behavioral is highly subjective, as the objective criteria that are applied to evaluate the model simulations and the thresholds for these objective criteria that define a simulation as “good” or “insufficient” are individual decisions. The decisions made in the presented work are listed on p.10L6-10.
We agree, that a different definition of a behavioral parameter set would affect the influence that the model parametrization has on the analyzed model outputs. The effect of the assumptions made in such an impact assessment were therefore discussed in section 4.2. Thus, the low number of behavioral parameter sets does not per se affect the sensitivity of the model outputs on the model parametrization.
The low number of behavioral parameter sets in this specific case results from the study design. A model parameter set was considered as a behavioral parameter set, if the

simulations performed with **all** model setups using that parameter set fulfilled the applied objective criteria (which means a model parameter set implemented in the Raab case study had to meet the thresholds for the objective criteria in all six implemented model setups).
This design decision may influence the resulting sensitivities of the model outputs, as the impact of the model setup and the model parametrization combined can be greater than the effect resulting from the illustrated study design (All individual model setups together with their model parametrizations may result in behavioral model/parameter setups, that are not considered here). The implemented setup however, isolates the effect of the model setups and the effect of the model parametrizations for model setups that were calibrated for a reference period and are applied for future changing conditions. In the context of an environmental impact study, to assess their individual effects is highly relevant in our opinion.
As a result of your comment (and similar comments in other reviews) we see a requirement to clarify the the evaluation of the parameter sets on p.10L6ff.

The revised text that should further clarify this issue is given on p.10L33 - p.11L14.

Figure 4: It is very hard to understand this figure. In my understanding the results from Fig. 3 are shown again and in addition to that the variations evoked by changes in land use or point emissions. Is it maybe better to present this as relative change to the lines in Figure 3? Or only as line and not as coloured area?

The Figures 4 to 8 follow the same pattern in the analysis that they illustrate. The Figures 4 to 8 indeed present the results shown in Figure 2 in a modified way. While Figure 3 shows the uncertainty bands of the 7000 simulations performed in each case study implementing the different combinations of input scenarios and model setups, the following figures separate the resulting uncertainty bands with respect to the discrete realizations of the individual model inputs and model setups. During the compilation of the manuscript we tested different ways to communicate the information we wanted to convey (e.g. plot all 7000 simulations as lines, analyze relative changes, etc.). We concluded however that the selected visualizations were the most suitable ones that supported our findings best. Thus, we prefer to remain with the presented figures. We see however the need to clarify the explanations regarding the Figures 3 to 8.

Section 3.2 was substantially modified.

Discussion: One idea is to add a table or figure as an overview in the discussion to show which of the five criteria has a dominant impact on discharge and nitrate and which criteria are uncertain. I think that the article would benefit from a clear and easy understandable presentation at the end as a kind of take-home-message. I have in mind a figure which summarize all results in relative values. To understand the overall idea of summary figures see for example Figure 9 in Herman et al., 2013.

We highly appreciate this comment and thank you for the link to the publication by Herman et al. (2013). Herman et al. (2013) used the summary figure as a very effective tool to summarize their findings. We were discussing how to implement this tool to summarize our findings in the manuscript. So far however, we were not able to come up with a good solution that would add value to the manuscript and facilitate interpretation for the reader.

Thus, unless we come up with an appropriate illustration during the revision of the manuscript, we prefer to not add an additional figure.

Specific comments

P.1, L. 8: I suggest to modify to: "In impacts studies in two Austrian catchments, …

We prefer your suggestion over the phrase in the manuscript. The text will be changed accordingly.

The text was changed accordingly.

P. 1, L.13: I suggest to write "for each catchments" instead of "for both catchments".

Together with other changes the section p.L11-14 will be updated as follows:
We developed scenarios of future changes for land use, point source emissions, and climate. The developed input ~~and implemented the~~ scenarios were implemented in ~~realizations in the different~~ SWAT model setups with different spatial aggregations and employing different model parametrizations that were able to adequately reproduce historical observations of discharge and $NO_3^- - N$ loads. ~~, which resulted in~~ In total 7000 combinations of scenarios and model setups were used to ~~both catchments. With all model combinations we~~ simulate~~d~~ daily discharge and $NO_3^- - N$ loads at the catchment outlets of each catchment.

The text passage was implemented as stated.

P.2, L: 17: I suggest to add: "using a set of different climate input data for hydrological models" at the end of this sentence (or a similar statement).

In this particular section we wanted to keep the statement more general (the statement is also true for land use change, or any other change process expressed with discrete scenarios). Thus, we prefer to keep the general phrase with the following example of climate change scenarios, as written.

P.2, L.27: The discussion on equifinality is not well motivated. I miss a sentence to relate both paragraphs.

We suggest the following modification of this section in the updated version of the manuscript:
To simulate the development of hydrological variables under changing conditions, the developed scenarios are implemented in hydrological models that are calibrated for historic conditions. Yet, often different model setups and different sets of parameters in a model can perform equally well to reproduce historical observations of the variables of interest. Equifinality is a well-known issue in hydrologic modeling that has been extensively addressed in the literature…

We added a paragraph on p.2L29-L32.

P.3, L. 5: I suggest to add a sentence at the beginning of the paragraph similar to "Sensitivity analysis can be used to derive the impact of different input variables on hydrological target variables" to make clear why you have selected this method.

The sentence will be added accordingly.

The text was changed accordingly.

P. 6, L. 11: fertilizer

This will be corrected accordingly.

The text was changed accordingly.

P. 8, L. 5: Please avoid one-sentence-paragraphs

The sentence will be added to the previous paragraph.

The text was changed accordingly.

P. 9, L: 14: I suggest to write: "applied a GSA on discharge and nitrate...".

To consider your suggestion and the suggestions made by other reviewers on this sentence it will be changed as follows:
In a ~~pre-analysis step~~ parameter screening, we applied a GSA to the simulations of discharge and $NO_3^- - N$ at the catchment outlets using all SWAT model setups ~~individually~~ to identify the relevant model parameters.

The text passage was implemented as stated.

Table 3: Is the sensitivity related to discharge or nitrate or both?

We appreciate this comment and think that this is valuable information for the reader. Thus, we suggest to modify Table 3 and differentiate between parameters that were influential for discharge related processes and $NO_3^- - N$ related processes.

Table 3 was modified accordingly. Due to its larger size in the updated version it was moved to the appendix (now Table A1).

P.15, L. 14: You may add that this result could be expected since the model structure is known to be of higher importance for low flows since high flows are strongly driven by the precipitation (observations).

We addressed this issue already in the discussion to some extent. We suggest however to stress this issue more and to clarify this point in the discussion. In contrast to your statement, the study design (that tried to assess the individual effects of the model setup and the parametrization) clearly show that the model setup has a stronger influence on large and medium discharges, whereas the model parametrization greatly affects the low flow.

P. 15, L. 31-34: For me, it seems to be that in Fig. 3, spring is the dominant season in the upper left subplot.

We agree with your observation that spring is little more dominant in Schwechat. Thus, we will mention this fact as well.

The text passage was implemented as stated.

Figure 4: The legend needs to be explained in the figure caption.

We agree that the used abbreviations are not self explaining in this figure. Hence, we will add an explanation of the abbreviations in the figure caption.

The text was modified accordingly on p.16L32-L34..

P. 19, L. 5: Could you add in which subplot you can see this drastic change?

Actually, that finding is supported by all subplots. The anomalies in precipitation affect the simulated long term monthly averages of discharge and $NO_3^- - N$ loads as well as the different segments of the illustrated flow duration curves of both catchments.

P. 19, L. 11: Have you an explanation for this?

We think that the precipitation anomalies explain these findings, to a large extent. Increases in mean annual precipitation increase the discharge and $NO_3^- - N$ loads, while a simulated reduction of mean annual precipitation in the future results in a reduction of discharge and $NO_3^- - N$ loads. This explanation is, in our opinion, provided in the text of the manuscript. We suggest however to revise this section and try to specify the statement more precisely.

P. 25, L. 3: I suggest to add "The selection of" before "climate scenarios".

This will be changed accordingly.

The text was changed accordingly.

P. 26, section 4.2: You may add a statement similar to "This analysis shows again that a clear description of the selected scenarios is mandatory for impact studies."

We appreciate this comment and such a statement will be added to the text.

An additional statement was added on p.27L23-L24..

References

Herman, J.D.; Kollat, J.B.; Reed, P.M.; Wagener, T. (2013): From maps to movies: high resolution time-varying sensitivity analysis for spatially distributed watershed models, Hydrol. Earth Syst. Sci., 17, 5109-5125.

RC2: 'modelling of hydrology and nitrate export from catchment', Anonymous Referee #2

This manuscript by Schürz et al. gives a detailed sensitivity and uncertainty analysis for modelling of hydrology and nitrate export in two medium-size catchments. The sensitivity analysis is elaborated for three groups of input scenarios (land use, point sources, climate) and alternatives of model setup and model parameters. The uncertainty of the modelled flow and nitrate exports is done separately for these five model-specific groups, which enabled evaluations of their influence on the reliability of modelling outputs.

I like the study. It shows a well-designed example how to transparently present modelling results. The methods are sound, using contemporary approaches, and sufficiently described. The results are suitably visualized and a discussed, and support conclusions.

We would like to thank the Anonymous Referee #2 for their positive and supportive feedback on this manuscript. In the following, we addressed each comment made by Anonymous Referee #2. The initial comments made are printed in *serif, italic font*. Our replies to the comments are written in black, non serif font nd our suggestions to revise the manuscript according to a comment are highlighted with the colors blue for insertions and red for deletions. The reference to the final modification in the revised manuscript is given below each reply to a comment in green.

Major comments

From my view, more credibility can be given to the parametrization of model (which shows very high impact to simulated results and uncertainty) when the selected parameter values that were used in the uncertainty analysis are given, at least in the Appendix.

Based on this comment and comments made by other reviewers, we propose to add the following information to provide further detail on the model parameters used:
To show any clustering of model parameter values of the selected parameters and to identify any parameter interactions we add the following figure in the Appendix of the manuscript:

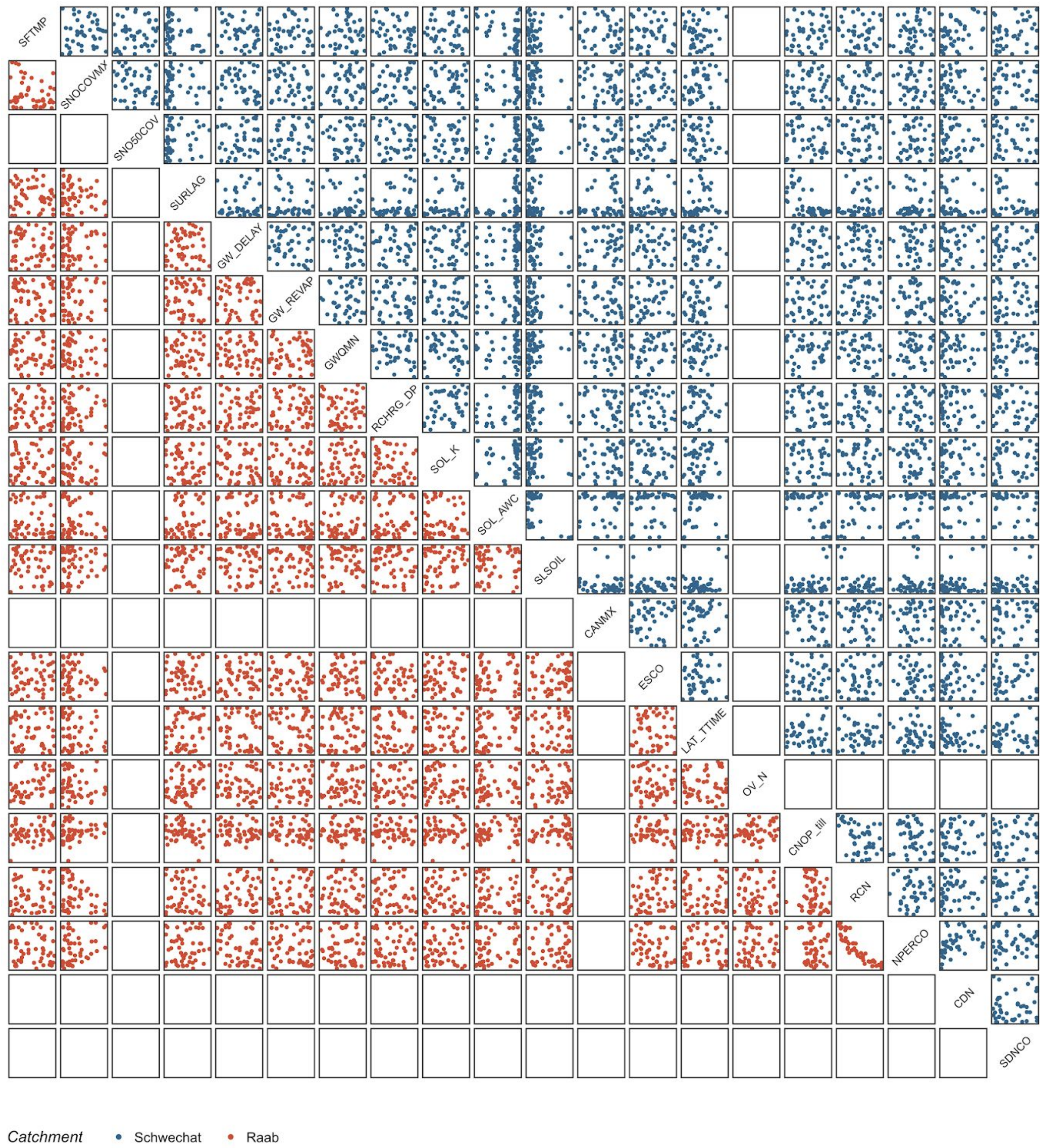

Figure caption:

Coordinate plot of the 43 and 52 behavioral SWAT model parameters that were used with the model setups of the Schwechat and the Raab, respectively. Each panel illustrates the connection of two model parameters for the Schwechat in red (below the diagonal) and the Raab in blue (above the diagonal). The x and y axes of each panel show the range of the respective parameter plotted along the x or y dimension. The corresponding parameter ranges for all illustrated parameters are provided in Table XX (Reference to table below).

Due to the limited space in the figure we avoided plotting any axes and axis labels. The figure however illustrates the clustering and interaction of model parameters. We additionally suggest to add parameter ranges and the type of change of the model parameters in an additional table:

Table caption:
SWAT model parameters calibrated in the model setups of the Schwechat and the Raab catchments. The type of change indicates whether the model parameters were replaced by absolute values, modified by adding absolute values to the predefined model parameters or, changed by a relative fraction of the predefined model parameter. Illustrated are the initial ranges of the model parameters and the ranges of the final behavioral parameter sets of the model setups of the Schwechat and the Raab catchments.

			Parameter change range	
Parameter	Type of change	Initial parameter range	Schwechat	Raab
SFTMP	replace value	[-1.00, 1.00]	[-0.69, 0.93]	[-0.98, 0.88]
SNOCOVMX	replace value	[100.0, 500.0]	[0.9, 177.0]	[100.8, 447.5]
SNO50COV	replace value	[0.20, 0.50]	[0.21, 0.49]	
SURLAG	replace value	[0.00, 0.50]	[0.02, 0.99]	[0, 0.1]
GW_DELAY	replace value	[0.0, 300.0]	[5.5, 25.0]	[2.1, 283.3]
GW_REVAP	replace value	[0.02, 0.20]	[0.05, 0.15]	[0.02, 0.20]
GWQMN	replace value	[0, 3000]	[566, 2472]	[109, 2925]
RCHRG_DP	replace value	[0.00, 1.00]	[0.31, 0.69]	[0.13, 0.97]
SOL_K	relative change	[-0.90, 10.00]	[0, 0.97]	[-0.79, 9.76]
SOL_AWC	relative change	[-0.90, 2.00]	[-0.86, 1.49]	[0, 1.98]
SLSOIL	replace value	[0.0, 150.0]	[0.9, 27.6]	[14.7, 148.2]
CANMX	relative change	[0.00, 0.25]	[0.34, 2.40]	
ESCO	replace value	[0.00, 0.90]	[0.05, 0.90]	[0.05, 0.89]
LAT_TTIME	replace value	[0.0, 180.0]	[0.8, 6.8]	[5.5, 176.3]
OV_N	absolute change	[-0.09, 0.60]		[0.07, 0.58]
CNOP_till	relative change	[-0.20, 0.10]	[-0.29, -0.06]	[-0.18, 0.01]
RCN	replace value	[2.00, 10.00]	[5.05, 9.97]	[2.3, 8.45]
NPERCO	replace value	[0.00, 1.00]	[0.24, 0.99]	[0.18, 0.7]
CDN	replace value	[0.00, 1.50]	[0.01, 1.44]	
SDNCO	replace value	[0.00, 0.50]	[0.02, 0.49]	

We provided further information concerning the model parametrization. The suggested figure was added as Fig. A3. The table providing the corresponding ranges of parameter changes as well as the initial parameter ranges are available in Table A2.

Specific comments

p.5, l. 25: Shouldn't be the Raab catchment area 988 km2?

Thank you for identifying that typo. According to Table A2 p.32 the total delineated area of the Raab catchment is 98815.9 ha. The value in the text on p.5 L25 will be changed accordingly from ~~998 km2~~ to 988 km2.

The text was changed accordingly.

p.19, l. 12-13: I suggest to join the sentences: "While a grouping of the individual climate scenarios with respect to their temperature anomalies shows a more indefinite picture, all climate scenarios simulated an increase in temperature."

This will be changed accordingly.

The text was changed accordingly.

RC3: 'Review of the manuscript by Schürz et al.', Anonymous Referee #3

I provide my comments below according to the HESS review criteria. Given some of my major comments below, it does not seem necessary to provide a more detailed line by line annotation at this point.

We want to thank the Anonymous Referee #3 for their detailed review of the manuscript and the valuable comments made to improve the quality of the manuscript. In particular the critical comments on the methodology helped us to reassess the results and the conclusions drawn in this work. The comments made by Anonymous Referee #3 are printed in *serif, italic font* below. Our replies to the comments are written in black, non serif font and our suggestions to revise the manuscript according to a comment are highlighted with the colors blue for insertions and red for deletions. The reference to the final modification in the revised manuscript is given below each reply to a comment in green.

1. Does the paper address relevant scientific questions within the scope of HESS? Yes. Trying to quantify and attribute uncertainty from various sources in "eco-hydrological" modelling in the context of climate and environmental change.

We appreciate the positive feedback on the relevance of our manuscript.

2. Does the paper present novel concepts, ideas, tools, or data? I found the way figures 2 to 7 very informative. I particularly found figure 2 very appealing in presenting SA results.

We appreciate the positive feedback on the visualization of our findings.

3. Are substantial conclusions reached? Given some of the discussions provided on the methodology below in NO.4, I am not sure if we can say conclusions are substantial.

A detailed reply to the specific comments can be found below (4 a) and b)).

4. Are the scientific methods and assumptions valid and clearly outlined? I very much liked how the manuscript tries to do a systematic and comprehensive approach, step-by-step, to set up the models, define scenarios, conduct SA/UA experiments, visualize (for better communication of) the results, and reach to conclusions. However, I have some major concerns about some of the methods and tools used in this study that I explain below:

a) Discrete PAWN SA: My most fundamental concern is related the way the main SA with PAWN is performed in this work, which also led to main conclusions in the paper. I strongly feel that the PAWN SA results (Figure 2) is largely impacted by the NUMBER of discrete realizations in each category (Table 4) and not by their CONTENT. In other words, it is intuitive that in this design of SA experiments, by default, the category with a higher number of members will always show a higher influence, because parameters sampled here will naturally have a much higher variability with respect to those categories. And this is exactly what we see in SA results and why some results are rather counter intuitive (e.g. negligible or small influence from land use changes or model setup, and very large influence from Climate and parameters). This is a fundamental issue that needs to be addressed by authors as it is the foundation for all conclusions.

We disagree with the argument that the influence of a model input / model setup depends on the number of realizations of that respective input/setup. Indeed, the model parametrization and the climate scenarios had the strongest impact on most of the analyzed processes and were represented by a substantially larger number of realizations compared to the other inputs. A counterexample to the statement that the sensitivity is per design impacted by the number of realizations of an input is illustrated by the influence of the point source scenarios in the Raab catchment for medium and low nitrate-nitrogen ($NO_3^- - N$) loads and $NO_3^- - N$ concentrations for medium and low discharges in this study. The calculated PAWN indices for these measures were substantially larger for the point sources compared to, for instance, the climate scenarios. Yet, only four point source scenarios were used, while 22 climate simulations were implemented.

We want to clearly point out however, that the number of discrete realizations of a model input can affect the calculation of a sensitivity index indirectly. In the case of the PAWN index a distance is calculated between the unconditional and the conditional cumulative distribution function (CDF) of a target variable (Pianosi and Wagener (2015) for example suggest to use the Komogorov-Smirnov test statistics). The unconditional CDF can also be estimated from all simulation that were performed (where all model inputs are perturbed), while to estimate the conditional CDF only simulations are used that used one discrete realization of the input of interest (this means all other inputs are perturbed, while the input of interest is kept constant). The distance measure is calculated for all realizations of a model input accordingly. The calculated distances for all conditional CDFs (keeping the model input constant at every respective realization) do have a certain distribution. To infer the PAWN sensitivity index, the calculated distances are summarized employing any summary statistics (Pianosi and Wagener (2015) for example suggests to use the median or the maximum). The choice of summary statistics can however also strongly affect the comparability of the calculated sensitivity indices of the individual model inputs if the distance measure distributions for the model inputs substantially differ. As a consequence, we employed the maximum statistics in this study, as we were primarily interested in the maximum possible influence an input has on an analyzed target variable. Different summary statistics, but also different methods for global sensitivity analysis (GSA, e.g. the method of Sobol (1993) that analyzes an average influence of a model input) were tested and evaluated during the compilation of this study. The outlined effects were observed in these analyses (yet not shown in this manuscript).

Further, the calculated sensitivities are well supported by the analysis of the simulation uncertainties. Inputs that showed a large influence on an analyzed process also showed a strong effect on the simulation uncertainty bands of that respective process.

Finally, we disagree with the statement that the negligible or small influence from land use changes or model setup are counter intuitive findings. In our opinion these findings were substantially discussed in section 4.1. Other literature cited in section 4.1 strongly supports the findings (e.g. Wagner et al. (2017), Guse et al. (2015), Mehdi et al. (2015a, 2015b), or Bieger et al. (2013) for the impact of land use change, or Jha (2014) for the model setups).

We gave this issue further attention and added the section on p.13L17 - p.14L.2

b) Design of Experiments: Authors do a great job particularly in explaining a rather careful and detailed procedure to setup the model, process the required data, define HRUs, and layout future land, pollution, and climate scenarios. This is extensive amount of work. However, I feel that this breadth has caused insufficient scientific depth in places in the manuscript. For example, it is unclear to me why certain various metrics are chosen in the SA analysis with VARS? How are these metrics really different from each other from an SA perspective (in particular, NSE and RSR are directly related, so why both are used?), Why this choice is not consistent with the metrics used in the next steps (e.g. what happened to KGE or RSR)? Perhaps strategically reducing some of the metrics can help in a more efficient way of conducting SA and presenting its results (e.g. some of the quantile classes presented in Figure 2 in each signature measure can be removed).

It is correct that different measures were used as objective criteria in the GSA to identify influential model parameters and in the model calibration (identification of behavioral parameter combinations). The purpose of the GSA was to screen the model parameters. This screening had an inclusive character, which means that the parameter had to be influential for at least one of the selected criteria. Consequently, the similarity of criteria did not affect the results of the parameter screening (if the measures are similar then the same parameters are influential for these objective criteria.). Contrary, the selection of behavioral parameters was exclusive. Thus, only criteria were used that describe the aspects of a simulated time series that we explicitly wanted to evaluate. In the selection of the criteria for the model calibration we referred to literature such as Pfannerstill et al. (2014).

We agree that the measures NSE and RSR are strongly related in their calculation. Yet, both measures differed completely in their application in this study. While the NSE was applied to the simulated and observed times series of a variable, the RSR was applied to various segments of the flow duration curves (FDC). Thus, the resulting NSE values also accounted for the timing of simulated values of a variable, whereas the RSR values of the FDC segments did only account for the distribution of simulated values of a variable letting aside the temporal occurrence of a value.

We fully understand that the Fig. 2 can overwhelm the reader, as we try to present a lot of information in one figure. Nevertheless, we think that all segments of a FDC characterize different processes of the water or the nutrient cycles (in this case). Further, the large number of analyzed segments of the FDCs visually support the gradual shifts of sensitivities of a target variable from one model input to others. There is a chance that this information is lost, when removing too many of the analyzed FDC segments from the figure.

To provide further detail for the STAR VARS method we added the section p10L6-L17. In general we strongly modified the section 2.6.1.

Or for example, what is the scientific reference or justification for the way UA is conducted here at the end using 7K simulations out of all possible combinations? Wouldn't a Latin Hypercube Sampling be a more effective choice than random sampling? These methods and choices (and other similar ones) must to be clearly justified in the manuscript.

As briefly mentioned above, other methods for GSA were tested as well (while not shown in the manuscript). A preceding analysis employed the Sobol method (Sobol, 1993) for GSA using a sampling design proposed by Saltelli (2002) that requires $N(k+2)$ samples, where N is the "base sample" (Saltelli, 2008) that was defined with 1000 in this study and k is the number of inputs (in this case 5).
As we identified issues with the average sensitivity that is expressed by the sensitivities calculated using the Sobol method (see also the reply 4a)) we utilized the random sample that was drawn for the Sobol method to calculate PAWN indices. Pianosi and Wagener (2018) outline how to estimate PAWN indices from any generic sampling. For this study the proposed concept was applied to discrete model inputs in this study.
We see however from this and other reviews on that matter, that the sampling and the confidence in the GSA results require greater attention in the manuscript. Thus, we suggest to revise the section of the input factor sampling in the revised version of the manuscript. Further, as proposed by Francesca Pianosi in her review, we plan to perform a bootstrapping (as presented in Sarrazin et al. (2016)) to calculate confidence intervals for the PAWN indices. This will greatly improve the results of the manuscript.

An explanation for the size of the sample was added on p.12L29-L32.

5. Are the results sufficient to support the interpretations and conclusions? Please see my comments above in NO.4.

We tried to clarify issues raised concerning the methodology that was applied to derive the results illustrated in the manuscript in 4 a) and b). Please find our replies to these comments below the respective sections 4 a) and b).

6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? No. Details of SA/UA experiments are missing. In particular, I found description of the VARS method somewhat short and there are important details that are missing (a more careful description from the original papers or some of newer applications is recommended). Another very important information that is missing is the ranges used for parameters, and an explanation of how these ranges are determined. These ranges can impact all the SA/UA results. Or it is unclear how parameters are tied to HRUs, and how all different setups, with different NO. of HRUs, in different basins have the same number of parameters (42) when doing SA with VARS?

We agree that the explanations concerning the parameter sensitivity analysis are rather short, as we intended to focus on the actual sensitivity study. Yet, you are right that the working steps in the parameter sensitivity analysis and the model parametrization affect the results of the following study.
We suggest to elaborate the parameter sensitivity analysis with greater detail. For the updated version of the manuscript adding a table is planned that provides information on the initial parameter boundaries, the boundaries of the final behavioral parameter sets and the type of change that was applied to the model parameters (whether the parameters were replaced by a single value globally or the spatially distributed parameter field was changed by a fraction of the parameter value or changed by adding/subtracting an absolute value).

The section 2.4 was strongly modified as a consequence. Table A2 and Fig. A3 were added to provide further detail to the model parametrization.

7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution? Yes.

Thank you

8. Does the title clearly reflect the contents of the paper? Yes for the most part.

Based on the assessment of the manuscript title we see no possibility to improve the title to more precisely reflect the contents of the manuscript.

9. Does the abstract provide a concise and complete summary? Yes.

Thank you

10. Is the overall presentation well-structured and clear? Yes for the most part.

Thank you

11. Is the language fluent and precise? I feel the language needs to be modified a bit. Both in terms of English grammar (double check usage of "the" and "comma"), and in terms of being scientifically more precise (e.g. using "pollution" instead of "emission"; or using "most influential input" instead of "most relevant"; or page 3 line 4; or page 3 line 26). I recommend a more careful review of the manuscript in this regard.

Thank you for the feedback on the language of the manuscript. Based on this comment and comments made by other reviewers we plan to carefully review the language in a revised version of the manuscript.

12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? Yes.

We appreciate your evaluation.

13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated? Some of the quantile classes presented in Figure 2 in each signature measure can be removed.

We outlined our thoughts on reducing the number of quantile classes in our reply on comment 4 a). Please see our reply above.

14. Are the number and quality of references appropriate? Yes.

We appreciate your evaluation.

15. Is the amount and quality of supplementary material appropriate? Yes.

We appreciate your evaluation.

References

Bieger, K., Hörmann, G., and Fohrer, N. (2013). The impact of land use change in the Xiangxi Catchment (China) on water balance and sediment transport, Regional Environmental Change, 15, 485–498.

Guse, B., Pfannerstill, M., and Fohrer, N. (2015). Dynamic Modelling of Land Use Change Impacts on Nitrate Loads in Rivers, Environmental Processes, 2, 575–592.

Mehdi, B., Lehner, B., Gombault, C., Michaud, A., Beaudin, I., Sottile, M.-F., and Blondlot, A. (2015a). Simulated impacts of climate change and agricultural land use change on surface water quality with and without adaptation management strategies, Agriculture, Ecosystems & Environment, 213, 47–60.

Mehdi, B., Ludwig, R., and Lehner, B. (2015b). Evaluating the impacts of climate change and crop land use change on streamflow, nitrates and phosphorus: A modeling study in Bavaria, Journal of Hydrology: Regional Studies, 4, 60–90.

Pfannerstill, M., Guse, B., and Fohrer, N. (2014). Smart low flow signature metrics for an improved overall performance evaluation of hydrological models, Journal of Hydrology, 510, 447–458.

Pianosi, F. and Wagener, T.(2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions, Environmental Modelling & Software, 67, 1–11.

Sarrazin, F., Pianosi, F., and Wagener, T. (2016) Global Sensitivity Analysis of environmental models: Convergence and validation, Environmental Modelling & Software, 79, 135–152.

Sobol, I. M. (1993). Sensitivity analysis for nonlinear mathematical models, Mathematical Modelling and Computational Experiments, 4, 407–414.

RC4: 'Review of the manuscript by Schürz et al.', Francesca Pianosi

The manuscript presents an interesting application of uncertainty and sensitivity analysis to the SWAT model. The aim is to assess the dominant controls of long-term discharge and nitrate-nitrogen load predictions under climate and land use change, while also taking into account the intrinsic uncertainty in the model, i.e. parameter and setup uncertainty. The analysis is solid and provides interesting insights about the model behaviour. Although the specific findings are only relevant to the investigated model and case studies, their discussion is interesting for the wider community of SWAT users and in general users of environmental impacts assessment models, as it demonstrates the type of findings yielded by GSA and their implications for the refinement and use of the model. The visual analysis introduced in Figure 4-8 is a simple and yet effective complement to quantitative GSA approaches.

Overall the paper is well structured and well written, and I think it should be accepted for publication.

Below are some points that could be addressed to improve the manuscript clarity before publication.

We would like to thank Francesca Pianosi for her constructive review and the valuable comments made. We appreciate the general positive feedback on the manuscript. It is a pleasure for us, that one of the developers of the PAWN sensitivity index evaluated this manuscript. Below, we addressed each comment made. We hope to clarify and discuss all points of concern sufficiently in the following document. The referee comments are printed in *serif, italic font*. Our replies to the comments are written in black, non serif font and our suggestions to revise the manuscript are highlighted with the colors blue for insertions and red for deletions. The reference to the final modification in the revised manuscript is given below each reply to a comment in green.

Major comments

[1] Language is at times unclear - some examples are given below as Minor points. I also have a general comment about the use of the term "sensitive". The authors use it as interchangeable with "influential" however I find this confusing, because "sensitivity" is an attribute of the output, not of the inputs. I would say that "input x1 is influential on the output" or "the output is sensitive to input x1" but I would not say that "input x1 is sensitive" - this is confusing. Some examples of these unclear occurrences are also given below under Minor points, however if the authors accept my remark they should check the entire manuscript.

We highly appreciate this comment and agree that the example provided above identifies the correct use of the terminology. We accept the remark and will improve the updated version of the manuscript to correctly use the terms “sensitive” and “influental” accordingly.

The terminology was revised in the entire manuscript.

[2] The definition and use of the behavioural parameter sets is slightly unclear. I think the confusion started on P. 10 L. 6-7 with the sentence "For all SWAT model setups of the Schwechat and the Raab catchments we identified non-unique parameter sets that adequately simulated daily observation of discharge and NO3-N loads".

Does it mean that you identified one behavioural parameter set for ***each*** *model setup, or that you identified one behavioural parameter set to be applied* ***in all*** *the setups? If the former, then how is the dependency between parameterisations and model setups accounted for in the GSA? If the latter, then the underlying assumption is that the same parameter values can effectively represent processes at different aggregation scales (ie for different definitions of the subbasins and HRUs)? This should be clarified. On a parallel note, I find it interesting that out of 100,000 sampled parameterisations only 43 and 52 where found behavioural. This is not uncommon in calibration of complex hydrological models but still worth highlighting. It would also be interesting to see whether these behavioural parameterisations are clustered in specific regions of the parameter space or if they are scattered across the sampled ranges, which would indicate a certain amount of interactions between the parameters. This could be illustrated for example through a parallel coordinate plot.*

We agree that the paragraph as stated leaves room for an ambiguous interpretation of the performed simulation and analysis steps. We designed the study the following way: For the different model setups of the Raab (6 different setups) and the Schwechat (4 different setups) we analyzed the model parameter sensitivities employing global sensitivity analysis (GSA). Thus, we performed six individual GSAs for the Raab catchment and four for the Schwechat catchment. The individual parameter sensitivity analyses resulted in the same sets of influential model parameters for the Raab catchment and the Schwechat catchment, respectively. As a consequence, we selected the same model parameters for all model setups of the Raab catchment and for all setups of the Schwechat catchment. For each case study, we drew 100 000 realizations of parameter combinations from the influential sets of model parameters. The simulations were performed with all model setups involving all drawn parameter combinations.
To answer your first question, a parameter set was eventually considered as a behavioral parameter set, if the simulations performed with **all** model setups involving that parameter set fulfilled the applied objective criteria stated on p.10L7-10.
This design decision was necessary, to treat the model setups and model parametrizations as individual inputs in the GSA (as you have indicated in the first part of your question). We agree with your comment that the selected layout implies the assumption that a parameter combination represents the analyzed processes at different aggregation scales. The individual models (i.e. a model setup with a specific spatial aggregation together with a selected model parametrization) were not analyzed and compared at the subasin or the HRU level. Yet, all models are capable of adequately simulating discharge and nitrate-nitrogen loads at the catchment outlets in the reference period. A drawback of this design decision is that it does not consider parameter combinations that that would result in satisfactory simulations when employed in one or several model setups but do not give good results with **all** model setups. Thus, the influence of the model setup and the model parametrization combined can be greater than the effect resulting from the illustrated study design. Nevertheless, we think that the presented results reveal relevant insights in their

current form, as they isolate the effect of the model setups that were calibrated for a reference period and are now applied for future changing conditions (the same applies to the parametrizations).
The selected experimental design is also a major reason for the low number of resulting behavioral parameter sets in the two case studies.
All of the explanations outlined above are in our opinion not sufficiently addressed in the current version of the manuscript. In particular, we think that the design decisions we made have to be conveyed and highlighted clearly in the methodology. Thus, we suggest to revise the concerning sections p.9L9 to p.11L2 accordingly.

The revised text that should further clarify this issue is given on p.10L33 - p.11L14.

According to the suggested visualization of the model parameters in a coordinate plot, we propose to add the figure below in the Appendix of the manuscript. We omitted any axes and tick labels due to very limited plotting space and the large number of parameters to visualize. The figure however illustrates any clustering and interaction of model parameters. We additionally suggest to add parameter ranges and the type of change of the model parameters in Table 3 on page 10.

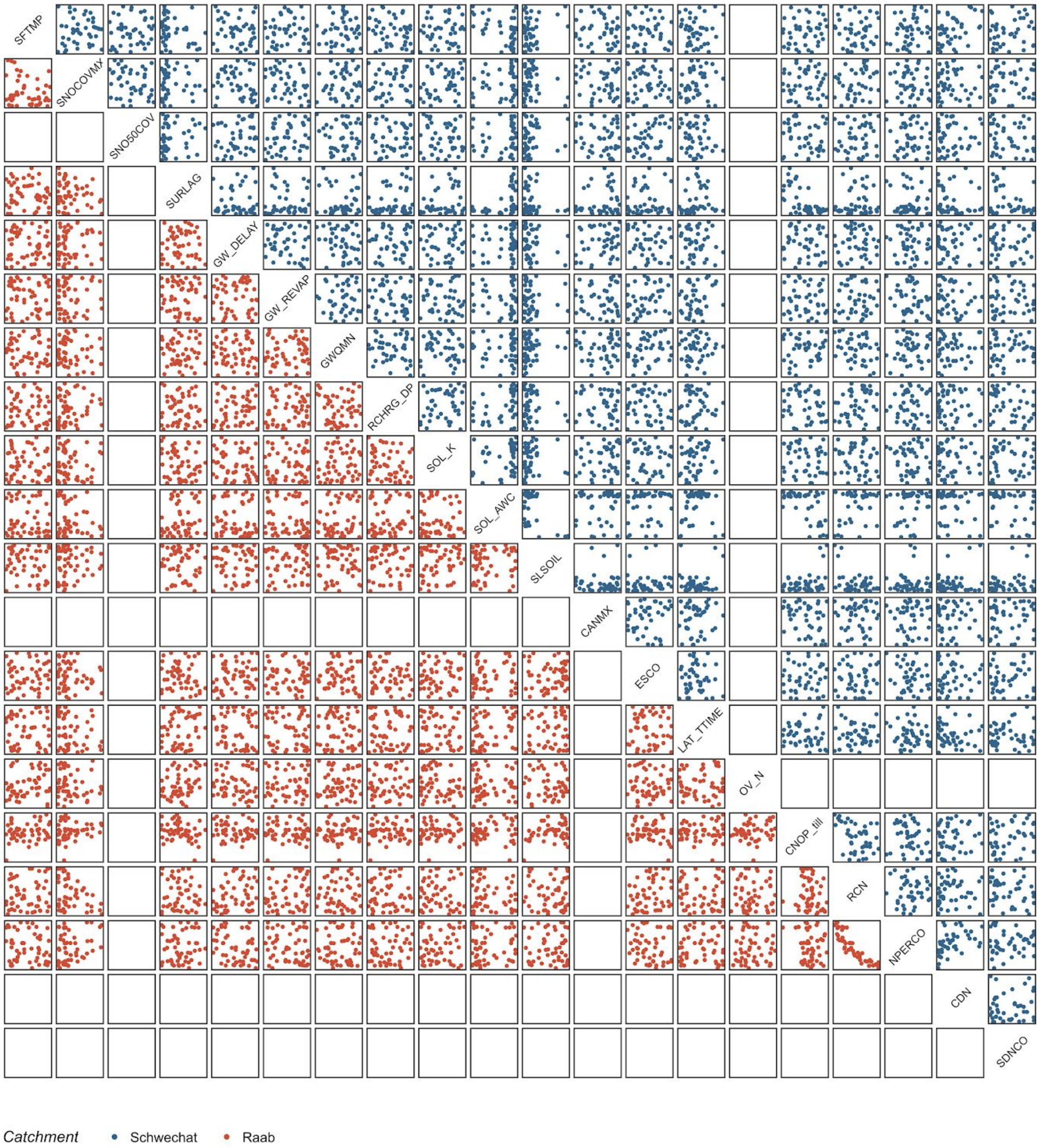

Figure caption:
Coordinate plot of the 43 and 52 behavioral SWAT model parameters that were used with the model setups of the Schwechat and the Raab respectively. Each panel illustrates the connection of two model parameters for the Schwechat in red (below the diagonal) and the Raab in blue (above the diagonal). The x and y axes of each panel show the range of the respective parameter plotted along the x or y dimension. The corresponding parameter ranges for all illustrated parameters are provided in Table XX (Reference to table below).

Table caption:
SWAT model parameters calibrated in the model setups of the Schwechat and the Raab catchments. The type of change indicates whether the model parameters were replaced by absolute values, modified by adding absolute values to the predefined model parameters or, changed by a relative fraction of the predefined model parameter. Illustrated are the initial ranges of the model parameters and the ranges of the final behavioral parameter sets of the model setups of the Schwechat and the Raab catchments.

			Parameter change range	
Parameter	Type of change	Initial parameter range	Schwechat	Raab
SFTMP	replace value	[-1.00, 1.00]	[-0.69, 0.93]	[-0.98, 0.88]
SNOCOVMX	replace value	[100.0, 500.0]	[0.9, 177.0]	[100.8, 447.5]
SNO50COV	replace value	[0.20, 0.50]	[0.21, 0.49]	
SURLAG	replace value	[0.00, 0.50]	[0.02, 0.99]	[0, 0.1]
GW_DELAY	replace value	[0.0, 300.0]	[5.5, 25.0]	[2.1, 283.3]
GW_REVAP	replace value	[0.02, 0.20]	[0.05, 0.15]	[0.02, 0.20]
GWQMN	replace value	[0, 3000]	[566, 2472]	[109, 2925]
RCHRG_DP	replace value	[0.00, 1.00]	[0.31, 0.69]	[0.13, 0.97]
SOL_K	relative change	[-0.90, 10.00]	[0, 0.97]	[-0.79, 9.76]
SOL_AWC	relative change	[-0.90, 2.00]	[-0.86, 1.49]	[0, 1.98]
SLSOIL	replace value	[0.0, 150.0]	[0.9, 27.6]	[14.7, 148.2]
CANMX	relative change	[0.00, 0.25]	[0.34, 2.40]	
ESCO	replace value	[0.00, 0.90]	[0.05, 0.90]	[0.05, 0.89]
LAT_TTIME	replace value	[0.0, 180.0]	[0.8, 6.8]	[5.5, 176.3]
OV_N	absolute change	[-0.09, 0.60]		[0.07, 0.58]
CNOP_till	relative change	[-0.20, 0.10]	[-0.29, -0.06]	[-0.18, 0.01]
RCN	replace value	[2.00, 10.00]	[5.05, 9.97]	[2.3, 8.45]
NPERCO	replace value	[0.00, 1.00]	[0.24, 0.99]	[0.18, 0.7]
CDN	replace value	[0.00, 1.50]	[0.01, 1.44]	
SDNCO	replace value	[0.00, 0.50]	[0.02, 0.49]	

The figure was added as Fig. A3 and the table was added as Table A2 as suggested.

[3] GSA was applied using 7000 samples of the input factors. How was this number chosen? Did the authors checked the adequacy of this sample size? The fact that the ranking based on the sensitivity indices in Figure 2 is confirmed by the visual analysis of Figure 4-8 is reassuring, yet formal methods exist to assess the robustness of the GSA results to the chosen sample size (for example, using bootstrapping confidence intervals as in Sarrazin et al. 2016 or a dummy parameter as in Zadeh et al 2017, both cited in the manuscript). It would be good to include more discussion of this point in the manuscript.

The number of input factor samples used in this study results from previous analyses performed using the present input factor data basis. A preceding analysis employed the Sobol method (Sobol, 1993) for global sensitivity analysis (GSA) using a sampling design proposed by Saltelli (2002) that requires $N(k+2)$ samples, where N is the “base sample” (Saltelli, 2008) that was defined with 1000 in this study and k is the number of inputs (in this case 5).

Similar to the generic sampling strategy that you suggested in a presentation at the EGU 2018 (Pianosi and Wagener, 2018a, that is now published in Pianosi and Wagener, 2018), we utilized a model input sample that was initially drawn for a GSA applying a different method (in our case the Sobol method) and employed it to estimate PAWN indices.
Besides the PAWN Index that is presented in this manuscript, we also tested the Sobol method and a modified version of the STAR-VARS method (Razavi and Gupta, 2016a, 2016b) in the course of this work (results for the latter two analyses are not shown in the manuscript). All experiments expressing the maximum sensitivity of the used target variables to the analyzed input factors showed strongly overlapping results. Thus, we were confident regarding the soundness of the GSA results. We agree however that it would be beneficial to the reader of the manuscript to provide any measure of confidence with the results of the GSA. Thus we suggest to perform bootstrapping (as demonstrated in Sarrazin et al. (2016)) to provide confidence intervals together with the calculated PAWN indices in section 3.2. p.15ff.

An explanation for the size of the sample was added on p.12L29-L32. As suggested, bootstrapping was performed. The methodology was added in section 2.6.1. The updated Fig. 2 now includes the results of the bootstrapping.

[4] The PAWN method was applied using a sampling scheme different from the one originally presented in Pianosi and Wagener (2015), in order to handle discrete-valued input factors. I understand the idea is to consider as fixed points x_i^j all the possible values that the discrete input factor x_i can take. Hence, for each input factor, the number of fixed points coincides with the number of possible values (n_i) that the input can take. If my interpretation is correct, then the text is misleading when it says (P. 13 L. 28) that "a generic random sample of the size N was drawn and subsetted with N/n_i subsets for all x_i^j" as the generic sample is divided into n_i (and not N/n_i) subsets. Is this right?

The term "generic" was used in the present context, as the sampling in all input factor dimensions was done randomly, although restricted by the number of fixed values each input can have. The separation of the total sample into N/n_i subsamples is then a required step to calculate the Kolmogorov-Smirnov distance for the input factor x_i at each location x_i^j. We understand however, that the term "generic sample" might be interpreted as a random sampling of continuous variables that is not the case here. Thus, instead of using the term "generic random sample" we suggest to specify the performed sampling in the following way (p.13L22-25):

~~The sampling scheme of PAWN index (Pianosi and Wagener, 2015) was initially introduced designed for continuous model parameters, which requires a modification for discrete model inputs.~~ Pianosi and Wagener (2015) introduced the PAWN sensitivity method using a specifically tailored sampling design to infer the PAWN indices T_i for continuous model inputs x_i. The ~~initial~~ proposed sampling strategy suggests to draw~~n~~ N_c conditional samples at n randomly sampled points of each influencing variable x_i, where xi is fixed at a value $x_i = x_i^j$ while all others are perturbed. Recently, Pianosi and Wagener (2018) extended the applicability of the PAWN sensitivity method to estimate T_i from a generic random sample of continuous model inputs. To approximate T_i the generic sample N is split into n segments along each model input dimension resulting in conditional samples N_c with an approximate size of N/n. We employed the proposed updated sampling strategy and adapted it for the use with discrete model inputs. ~~Instead, a generic random~~ A sample of the size N was drawn. For each model input combination every model input was sampled randomly from its discrete realizations. ~~and subsetted with~~ N/n_i ~~subsets for all~~ x_i^j ~~t~~ To infer ~~the values for all~~ $KS_j(x_i)$ for all discrete values x_i^j of a model input x_i the sample N was split into subsets for all n_i discrete values, resulting in subsets of the size N/n_i on average. It is important to consider, that the subset size depends on the number of discrete values n_i of a model input x_i, while the subsets of the sampling scheme proposed by Pianosi and Wagener (2018) were on average N/n for all model inputs x_i.

The text section was modified as suggested.

Also, if I understand the strategy correctly, then the inputs with small number of possible values (for instance the land use scenario) are associated with conditional distributions based on a very large number of samples (around N/n_i =7,500/2 in the case of land use scenarios), while the inputs with large number of possible values (for instance the parameterisation) are associated with conditional distributions based on much smaller number of samples (around 7,000/43). Do you think using such different sample sizes could have had an impact on the estimation of the KS values and hence of the PAWN sensitivity indices?

Your assumption concerning the subset sizes is correct. We admit that the manuscript does not convey this information clearly. Thus, we suggest to update this section of the manuscript as proposed in the reply above. In the current version of the manuscript we did not analyze the effect of the strong differences in the subset sizes on the confidence intervals of the calculated sensitivities. We did however compare the results derived with the PAWN method to the results inferred from an adapted version of the STAR-VARS method (Razavi and Gupta; 2016a, 2016b) that was not affected by the different numbers of discrete values for each model input due to its sampling design. We observed only minor differences between these methods and hence assumed that the effect of the different subset sizes is low. We suggest however for the updated version of the manuscript to consider that point in the bootstrapping. If that assumption is correct, we expect that if the impact of the different subset sizes is low when the confidence intervals remain in a comparable range for different numbers of subsets of the individual model inputs.

The updated results are shown in Fig.2. The bootstrapping were additionally addressed in the results on p.15L23-L29.

Finally, a new sampling strategy was recently proposed for PAWN (Pianosi and Wagener, 2018). While this new strategy is still designed for continuous inputs, and hence could not be used here, it would be good to mention its existence for readers who may want to apply PAWN in the future (as for the case of continuous inputs this would be recommended over the strategy in the 2015 paper).

The publication will be considered in the updated version of the manuscript, as suggested in the updated section above.

The reference was added in the revised version of the manuscript (e.g. in section p.14L4-L17).

[5] I think the discussion in Section 4.2 is interesting but potentially slightly misleading. The authors clarify that "several assumptions were made in the development of scenarios that are highly subjective". I understand the importance of highlighting the subjectivity inherent in the scenario definition if the goal of this study was to make projections of the future evolution of the two catchments. However, this is not the objective when doing GSA. GSA answers the question: "how much output variation do we get if we vary the inputs within certain ranges?" The answer yielded by GSA (i.e. the sensitivity indices, the input ranking, etc.) is certainly conditioned upon the chosen ranges, however this is "intrinsic" to the question asked, regardless of how the choice is made - be it an "objective" calibration exercise (as done for the parameterisations) or a "storylines" approach. In other words, I think the point is to justify why certain scenarios are considered for the impacts assessment study; once they have been selected for that purpose, it follows that they would be used in the GSA too if one wants to know their relative influence with respect to other input factors of the model. So, I do not agree with the sentence (P. 26 L. 13-14) "For the SA of the simulated variables the diversity of the developed scenarios is essential.": diversity may be important for the impacts assessment (is it?) but not necessarily for the GSA. If a limited set of scenarios were selected for the impacts assessment, I would use that set for the GSA even if it is not diverse.

We agree with the comment, that the goal of a GSA is to attribute the variations in simulated outputs to variations in model inputs, rather than simulating possible futures for a catchment. This was not the message we wanted to convey with this study. We intended to point out how GSA and an analysis of the uncertainty bands as illustrated can complement any impact study in understanding the sources of the uncertainties in simulating future conditions.
Maybe this is again, an issue of terminology. What we specifically wanted to address in this section was the fact that the subjective decisions we as modelers make in developing future scenarios will, no doubt, affect the simulation of a variable of interest. Further, when the developed discrete scenarios for a model input result in a wide range of a simulated output this will also affect the sensitivity of the output variable to the respective model input.
The analogue example for a single continuous model parameter would be to change the interval of that parameter in which it can vary for an assessment of the its influence on the model output. The selection of the parameter range is apparently highly subjective as well. Yet, while increasing the interval of a continuous property to cover more extreme regions of the model input space is a simple concept, the impact on the simulation of an output variable caused by any assumptions made in the development of model input scenarios is not always entirely clear in the scenario development. As this issue is not always addressed appropriately in environmental impact studies (e.g. by only using a few climate scenarios in an impact assessment), we saw a high need for this important discussion.
With the term “diverse” we wanted to express to represent a wide range of possible future representations of a model input. The addressed sentence seems however to be redundant as the following sentence repeats the argument. Thus, we suggest to change this section as follows:
~~For the SA of the simulated variables the diversity of the developed scenarios is essential. Thus, s~~ Scenarios must cover a broad range of possible futures...

The text was changed accordingly.

Minor points

P. 1 L. 15: "scenario inputs" should be "input scenarios"

This will be changed accordingly in the updated version of the manuscript.

The text was changed accordingly.

P. 2 L. 5: "the precipitation of the climate scenarios" sounds a bit odd, maybe "precipitation projections"

We agree with this comment and suggest to change the sentence as follows:
Additionally, the visual analysis of the uncertainty bands illustrated that the anomalies in precipitation of the different climate scenarios dominated the changes in simulation outputs, ~~rather than changes~~ while the differences in air temperature in both case studies showed no considerable impact.

The section was modified as given now on p.2L6-L8.

P. 3 L. 3-4: "An assessment is only as good as the dominant contributors of uncertainty in such a modeling chain." Unclear. Something seems to be missing in this sentence: an environmental impacts assessment is only good if dominant contributors of uncertainty are... what? Identified? removed? ...?

The sentence actually does not contribute much information. Thus, we rather suggest to delete it in the updated version of the manuscript.

The text was deleted.

P. 3 L. 11-12: "model computations" should be "model evaluations" (or "runs" or "executions")

The phrase will be changed to: ... from certain a number of model ~~computations~~ evaluations.

The text was changed accordingly.

P. 3 L. 19: "Most applications utilize GSA to identify and rank continuous model parameters". Unclear: GSA does not "identify parameters" at most "identify influential parameters"

Will be changed to:
Most applications utilize GSA to identify influential model parameters and to rank ~~continuous~~ model parameters according to their influence on model outputs. Model parameters are usually continuous model inputs.

The updated section is now given on p.3L23-L25.

P. 3 L. 21: "Although," Comma should be removed

This will be changed accordingly.

The text was changed accordingly.

P. 3 L. 26-27: "An OAT analysis however presumes linear models and non-correlated inputs". Not sure OAT requires a linear model, for instance the Morris method uses a OAT approach and yet is typically applied to nonlinear models. More generally, why should GSA be applied to a linear model at all? If the model is linear than the effect of each input on the model output is simply proportional to the input variation, no need to do GSA to know that.

In this section we highlighted the equivalency of the standard procedure performed in impact assessments an **local** "one-at-a-time" (OAT) analyses. We agree that in this specific sentence we used the more general term "OAT" instead of referring specifically to "local OAT". The presumptions of OAT such as linearity of the model or independence of model inputs was also addressed by Baroni and Tarantola (2014) or Saltelli (2010). To infer the (global) sensitivity of a model output from a delta change of a model input presumes that the same delta change of the model input at another position in the input space has the same effect on the model output and is not influenced by any other model input (linearity and independence). We further did not suggest to apply sensitivity analysis to a model where a linear relationship between the inputs and outputs is a-priori known. Contrary to that, we state that applying an OAT analysis to infer sensitivities of model outputs implies model linearity and the independence of model inputs.
Concerning terminology we referred to the terms as they are used in Saltelli and Annoni (2010), where OAT was considered to be performed from the same "nominal point" and the analogous analysis performed at various points in the input space was termed "radial" elementary effects (EE). We do however agree that methods such as the Morris method or EE also employ OAT sampling designs while inferring global estimates of the output sensitivities.
Thus, we suggest to modify the commented sentence and add "local" to specifically address the issues with local OAT analysis.

The text was changed accordingly.

P. 4 L. 4: "complex". Unclear. What is the definition of a "complex" input?

The term "complex" was already used earlier in the manuscript (e.g. p.3L21, p.3L30). Yet, we do not provide an explanation of that term at any point in the manuscript. Further, the term "complex" apparently does not clearly convey what is meant here, where maybe "composite" might be a more precise term to use. Thus we suggest to change the term "complex" to "composite" and further add examples in p.3L21:
Although, it is possible to implement ~~more complex~~ composite model inputs (e.g. climate scenarios that affect several climate variables at the same time, or land use scenarios that can impact the model setup) in GSA…

The term 'complex' was replaced by 'composite'.

P. 4 L. 5: "No study is known to us that takes advantage of GSA in the scope of environmental impact studies." What is the definition of "environmental impact studies" here? I would say that GSA has been applied to such studies before, e.g. Anderson et al (2014); Butler et al (2014); Le Cozannet et al (2015)

We agree that if you consider "environmental impact studies" in their actual broad context that GSA has been applied in several studies. Thus, the sentence is misleading here and will be deleted. The publications mentioned here as examples should rather be acknowledged and mentioned in the introduction.

We considered the suggested literature and deleted the sentence accordingly.

P. 4 L. 13-16: Very long sentence, consider splitting into two.

Based on the GSA and the visual analysis of the simulated uncertainties we are able to draw conclusion on the simulation of discharge and $NO_3^- - N$ loads as impacted by the model setup, model parametrization and the future scenarios of land use, point source emissions and climate. These conclusions are of course limited to assumptions made in the model setup and in the development of the scenarios.

The sentence was modified as suggested.

P. 8 L. 3: "Although," Comma should be removed.

This will be changed accordingly.

The text was changed accordingly.

P. 8 L. 14-15 "The SWAT model setups for the Raab and the Schwechat involved decisions for the selected number of subbasins of a model setup and the definition of the HRUs." Convoluted sentence.

This section will be modified as follows:
~~The SWAT model setups for the Raab and the Schwechat involved decisions for the selected~~
~~number of subbasins of a model setup and the definition of the HRUs. Both modifications are necessary decisions for any SWAT model setup.~~
A SWAT model setup requires the modeler to determine an "appropriate" number of subbasins and to make decisions for the HRUs (such as eliminating "insignificantly" small HRUs from the setup). The SWAT model setups for the for the Raab and for the Schwechat had different numbers of subbasins and defined HRU differently.

This section was modified as suggested.

P. 8 L. 15: "Both modifications": which modifications? Unclear

Please see the changes suggested in the comment above.

P. 9 L. 2: "involving". Unclear. Maybe "which requires"?

The text will be changed accordingly.

The section 2.4. was strongly modified in general.

P. 9 L. 11: "to define of the thresholds". Remove "of"? In general, the entire sentence is a bit unclear. How is the "aggregation error" defined? Error in which variable, and with respect to what "correct" value?

Thank you for finding the typo. As this section requires further explanations this section will be modified as follows:

In total, we set up four SWAT models, two with 3 and two with 14 subbasins for the Schwechat catchment and six setups for the Raab catchments with two each of 4, 29, and 54 subbasins. We kept the resulting HRUs of full HRU setups unmodified. The numbers of HRUs in the reduced HRU setups were modified by applying thresholds for land use, soil, and slope classes. HRUs with an area below the defined thresholds were eliminated from a model setup. ~~To maintain a comparable aggregation error with the number of subbasins for the different model setups, w~~ We employed the R package topHRU (Strauch et al., 2016) to ~~define of the~~ determine optimum thresholds for land use, soil, and slope classes ~~and accepted~~ that result in a maximum aggregation error of 5% of the total area of the HRUs when comparing the changes of land use, soil, and slope classes of the full HRU setup and the reduced HRU setup with the same numbers of subbasins. Table 2 gives an overview of the final baseline model setups for both case studies.

The section 2.4. was strongly modified in general.

P. 9 L. 14: "In a pre-analysis step," In the GSA literature, this kind of "pre-analysis step" is often called a "screening" analysis, as it aims at screening out the non-influential parameters. Maybe worth mentioning the term as it would be familiar to many readers.

We appreciate the suggestion. The text will be changed accordingly.

The text was modified (see p.9L21).

P. 9 L. 14: "relevant parameters". Relevant to what? Maybe better "influential"

The term will be changed accordingly.

The text was changed accordingly.

P. 9 L. 21: "FDCs". Explain the acronym

The acronym “FDC” was introduced on p.4L11:

...as well as flow duration curves (FDCs) of daily discharge and daily $NO_3^- - N$ loads...

P. 13 L. 5: "To identify the impact of" maybe better "To measure the relative importance of"

We prefer your suggestion. Thus, we will implement it accordingly.

The text was changed accordingly.

P. 13 L. 7: "PAWN involves". Unclear what "involves mean. Maybe better "PAWN uses"

The text will be changed to PAWN employs...
The text was changed as suggested.

P. 13 L. 11: "the sensitivity of a model input x for a target variable y". Sensitivity is an attribute of the output, not the input. I would rephrase as "the sensitivity of a target variable y to a model input x".

As mentioned in our reply to the major comment [1] the entire manuscript will be modified to meet this suggestion. Thus, we also implement this suggestion.

P. 13 Eq. (1) and (2). The mathematical notation could be made clearer. I find it odd that in Eq. (1) KS takes as subscript the index of the fixed point (j) while its argument remains the generic input x_i. This choice also makes it more difficult to understand how maximisation occurs in Eq. (2). I think using the notation $KS(x_i^j)$ in both equations would make things much clearer.

We appreciate this comment and will change the equations accordingly.

The equations on p13 and p14 as well as the corresponding text were modified accordingly.

P. 13 L. 23: "possible states". Why "states"? The term was never used with this meaning before. I would rather say "possible values".

This is a remnant of a previous version of the manuscript and will be changed accordingly.

The text was changed accordingly.

P. 13 L. 24: "a lower sensitivity of the input xi on the target variable y". Again, rephrase as either "a lower sensitivity of the target variable y to the input x_i" or "a lower influence of the input x_i on the target variable y".

This suggestion will be implemented.

The text was changed accordingly.

P. 13 L. 28: "subsetted with" Not sure "subset" can be used as a verb. Maybe better "divided into"

Your statement is correct. The verb “subset” is not listed in any dictionary. Thus, we will rephrase it as suggested.

The text of section 2.61 was strongly modified. The verb ‘subset’ was replaced by other terms in the entire document.

P. 13 L. 29. "... were used for the sensitivity assessment". I would link this to the mathematical notation just introduced above and say: "... were used as target variable y".

We appreciate the suggestion and will implement it accordingly.

The text was changed accordingly.

P. 14 L. 10-12: "In this study, we consider all execute model setups to be plausible..." I do not understand this clarification. What other approach would have been possible? To discard some simulations because deemed not plausible? And how would you define then what is plausible and what is not? Please clarify.

We agree that the phrasing sounds odd. Thus, this sentence will be rephrased to:
~~In this study, we consider a~~ All executed model setups ~~to be~~ represent plausible realizations of the future conditions in both catchments to simulate future discharge and $NO_3^- - N$ loads.

The text was changed as proposed.

P. 14 L. 17: "low number of each input". Unclear. Do the authors mean "low number of inputs" (i.e. 5 inputs) or "low number of possible values taken by each input"?

Here we meant the latter. We will update that phrase to:
~~The low number of each input included in the study~~ The low number of possible values taken by each input allowed…

The text was changed as proposed.

P. 14 L. 24-28. This sounds like a repetition of what just said in the methodology section, I do not think is needed. I would rather use this opportunity to explain how to read

Figure 2 (what is the difference between the panels and how to read each circle plot).

We will consider to replace this section with an explanation of Figure 2 in the updated version of the manuscript. As further modifications will be added to the figure (e.g. confidence intervals, etc.) we do not suggest modifications to the text at this point.

The section was updated on p.15L16-L22 in the revised version of the manuscript.

P. 15 L. 17: "highly sensitive" replace by "highly influential"

This will be changed accordingly in the revised manuscript.

The text was changed accordingly.

P. 15 L. 25-26: "their overall sensitivities follow the general trend of the climate scenarios to a large extent". Unclear, please rephrase.

We will rephrase this section as follows:
~~For most of the analyzed signature measures t~~ The model setups yielded insignificantly low PAWN indices for the majority of signature measures with values below 0.1 in the Raab case study, indicating that the model setup ~~was not sensitive~~ had a low influence on most analyzed processes. ~~Although the Raab case study shows low sensitivities for the model setups, their overall sensitivities follow the general trend of the climate scenarios to a large extent.~~ The pattern of the resulting PAWN indices of the model setups closely follows however the pattern of the PAWN indices that were calculated for the climate scenarios.

The section was rephrased as proposed.

P. 15 L. 33: "difference that is visible for the two". Unclear, do you mean "difference that is visible between the two"?

This sentence will be rephrased to:

~~A substantial difference that is visible for~~ A notable difference between the two case studies is how ~~the reference period relates to the uncertainty bands in amplitude~~ the simulations of long term monthly discharges and $NO_3^- - N$ loads in the reference period compare to the ranges of future simulations.

The section was rephrased as proposed.

P. 15 L. 33-34: "how the reference period relates to the uncertainty bands in amplitude". Unclear what this means.

See our suggestion for the modification of that phrase above.

P. 16, caption of Fig. 2: "Model input sensitivities for signature measures". Replace by "sensitivities of signature measures to model inputs". And later on "sensitivities of" should be replaced by "sensitivities to"

The caption will be changed accordingly in the updated version of the manuscript.

The caption of Fig.2 was substantially modified also considering this suggestion.

P. 17, text and Figure 3: what does "specific discharge" mean? Why "specific"?

I am not sure how well established this term is used in the hydrologic community. The specific discharge relates a discharge (given in e.g. $m^3 s^{-1}$) to the catchment area that produces the runoff and sums it for a specific time interval (in this case on monthly basis). We decided to use the specific discharge for a better comparison of the two catchments with substantially different catchment sizes.

P. 17 L. 6: "show a difference". Does this mean "show an increase"? If so, I would use "increase", it makes it easier for the reader to follow.

We prefer your suggestion. This will be changed accordingly in the updated manuscript.

The text was changed accordingly.

P. 19 L. 12: "While a grouping...". Remove "While".

This will be changed accordingly.

The text was changed on p.19L20-21.

*P. 24 L. 20: caused future land use change" Maybe "caused **by** future..." ?*

This will be changed in the updated version of the manuscript.

The text was changed accordingly.

P. 26 L. 31-32: "The application of sampling strategies for SA usually do not account for the circumstances that one model input constrains any other model input". I do not fully agree. There is an increasing literature on GSA methods applicable to the case of dependent inputs, see for instance Mara and Tarantola (2012; 2017).

We appreciate the comment and I think it is worth to mention and acknowledge these publications. While it is a more straight forward approach to constrain a continuous property by another continuous property it might be not a straight forward procedure to identify all plausible scenario combinations for multiple model inputs (e.g. some future climate settings might make future agricultural practices implemented in a land use scenario impossible). In the context of the present work we were referring to the latter case. We suggest to clarify this in the manuscript and to acknowledge the substantial work that was done to constrain dependent continuous variables in GSA.

The section was updated on p.28L7-L10.

P. 27 L. 17-18: "by a factor of up to 5... up to 8". Do these numbers come out of a comparison of PAWN indices? If so, I am not sure I would draw such quantitative comparison. PAWN indices are (maximum) KS values: what is the practical interpretation of "a factor of 5" between KS values? I find it difficult to imagine.

We appreciate your feedback on that section. As a consequence we will remove that comparison in the updated version of the manuscript.

The text was deleted as suggested.

P. 28 L. 3: "the lack of tool that allow the practitioners access to such methods". Not sure I understand what the authors mean here. Several GSA software tools are available (some are reviewed for example in Pianosi et al 2015). So what is the problem here? That they are not "friendly" enough for practitioners to use them? Or that they are not sufficiently tailored to SWAT applications? Pls clarify.

We agree that the sentence is too vague. Software, toolboxes and libraries to perform GSA are available for many different programming languages, for instance the SAFE toolbox (Pianosi et al., 2015) for matlab, SPOTPY (Houska et al., 2015) for python, or R packages such as sensitivity (Iooss et al., 2015), or fast (Reusser, 2015).
From a practitioner's perspective the challenge is to assemble such a large number of models and to perform thousands of model simulations for a large number of model input combinations, instead of performing the status quo procedure of implementing single scenarios into a calibrated model. To generalize such analysis for the application in environmental impact studies we suggest to come up with frameworks that support the practitioner in this laborious working steps of a case study.
Thus we suggest to specify the section p.28L2-4 as follows:
The main constraint for a practical application however, remains the lack of tools that allow practitioners access to such methods. As a consequence, we plan to implement the

demonstrated procedures and tools for visualization into a user friendly programming environment.

The section was updated on p.29L14-L20.

References

Anderson, B., Borgonovo, E., Galeotti, M., Roson, R., 2014. Uncertainty in climate change modeling: can global sensitivity analysis be of help? Risk Anal. 34 (2).

Butler, M.P., Reed, P.M., Fisher-Vanden, K., Keller, K., Wagener, T., 2014. Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis. Environ. Model. Softw. 59.

Le Cozannet et al (2015) Evaluating uncertainties of future marine flooding occurrence as sea-level rises, Environmental Modelling Software, 73.

Mara and Tarantola, 2012, Variance-based sensitivity indices for models with dependent inputs, Reliability Engineering and System Safety

Tarantola and Mara, 2017, VARIANCE-BASED SENSITIVITY INDICES OF COMPUTER MODELS WITH DEPENDENT INPUTS: THE FOURIER AMPLITUDE SENSITIVITY TEST, International Journal for Uncertainty Quantification

Pianosi and Wagener (2018), Distribution-based sensitivity analysis from a generic input-output sample, Environmental Modelling Software

Pianosi, F., Sarrazin, F., Wagener, T. (2015), A Matlab toolbox for Global Sensitivity

Baroni, G. and Tarantola, S. (2014). A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study, Environmental Modelling & Software, 51, 26–34.

Houska, T., Kraft, P., Chamorro-Chavez, A., Breuer, L. (2015). SPOTting model parameters using a ready-made python package. PloS one, 10(12), e0145180.

Iooss, B., , Janon, A., Pujol, G., (with contributions from Boumhaout, K., Da Veiga, S., Delage, T., Fruth, J., Gilquin, L., Guillaume, J., Le Gratiet, L., Lemaitre, p., Nelson, B., Monari, F., Oomen, R., Rakovec, O., Ramos, B., Roustant, O., Song, E., Staum, J., Sueur, R., Touati T., and Weber, F. (2018). sensitivity: Global Sensitivity Analysis of Model Outputs. R package version 1.15.2. https://CRAN.R-project.org/package=sensitivity

Pianosi, F. and Wagener T. (2018b). A new implementation of the PAWN method to perform density-based sensitivity analysis from a generic sample, EGU General Assembly 2018, Geophysical Research Abstracts Vol. 20, EGU2018-10784.

Razavi, S. and Gupta, H. V. (2016a). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory, Water Resources Research, 52, 423–439.

Razavi, S. and Gupta, H. V. (2016b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application, Water Resources Research, 52, 440–455.

Reusser, D. (2015). fast: Implementation of the Fourier Amplitude Sensitivity Test (FAST). R package version 0.64. https://CRAN.R-project.org/package=fast

Saltelli, A. (2002). Making best use of model valuations to compute sensitivity indices. Computer Physics Communications 145, 280–297.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). Global Sensitivity Analysis. The Primer, John Wiley & Sons, Ltd, Chichester, UK.

Saltelli, A. and Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis, Environmental Modelling & Software, 25, 1508–1517

Sarrazin, F., Pianosi, F., and Wagener, T. (2016) Global Sensitivity Analysis of environmental models: Convergence and validation, Environmental Modelling & Software, 79, 135–152.

Sobol, I. M. (1993). Sensitivity analysis for nonlinear mathematical models, Mathematical Modelling and Computational Experiments, 4, 407–414.

A comprehensive sensitivity and uncertainty analysis for discharge and nitrate-nitrogen loads involving multiple discrete model inputs under future changing conditions

Christoph Schürz[1], Brigitta Hollosi[2], Christoph Matulla[2], Alexander Pressl[3], Thomas Ertl[3], Karsten Schulz[1], and Bano Mehdi[1,4]

[1]Institute of Water Management, Hydrology and Hydraulic Engineering, University of Natural Resources and Life Sciences, Vienna (BOKU), Vienna, Austria
[2]Department of Climate Research, Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienna, Austria
[3]Institute of Sanitary Engineering and Water Pollution Control, University of Natural Resources and Life Sciences, Vienna (BOKU), Vienna, Austria
[4]Division of Agronomy, Department of Crop Sciences, University of Natural Resources and Life Sciences, Vienna (BOKU), Tulln, Austria

Correspondence: Christoph Schürz (christoph.schuerz@boku.ac.at)

Abstract. Environmental modeling studies aim to infer the impacts on environmental variables that are caused by natural and human-induced changes in environmental systems. Changes in environmental systems are typically implemented as discrete scenarios in environmental models to simulate environmental variables under changing conditions. The scenario development of a model input usually involves several data sources and perhaps other models, that are potential sources of uncertainty. The setup and the parametrization of the implemented environmental model are additional sources of uncertainty for the simulation of environmental variables. Yet, to draw well-informed conclusions from the model simulations it is essential to identify the dominant sources of uncertainty.

In ~~two Austrian impact studies~~ impact studies in two Austrian catchments the eco-hydrological model Soil and Water Assessment Tool (SWAT) was applied to simulate discharge and nitrate-nitrogen (NO_3^--N) loads under future changing conditions. For both catchments the SWAT model was set up with different spatial aggregations~~and non-unique~~. Non-unique model parameter sets were identified that adequately reproduced observations of discharge and NO_3^--N loads. We developed scenarios of future changes for land use, point source emissions, and climate and implemented the scenario realizations in the different SWAT model setups with different model parametrizations, which resulted in 7000 combinations of scenarios and model setups for both catchments. With all model combinations we simulated daily discharge and NO_3^--N loads at the catchment outlets.

~~In~~ The analysis of the 7000 generated model combinations of both case studies ~~we employed global sensitivity analysis (GSA~~had two main goals; i) to identify the ~~impact of the scenario inputs~~ dominant controls on the simulation of discharge and NO_3^--N loads in the two case studies, and ii) to assess how the considered inputs control the simulation of discharge and NO_3^--N loads. To assess the impact of the input scenarios, the model setup and the parametrization on the simulation of discharge and NO_3^--N loads ~~. We accompanied the GSA with a visual analysis of the simulation outputs and their associated uncertainties~~ we employed methods of global sensitivity analysis (GSA). The uncertainties in the simulation of discharge and

NO_3^--N loads that resulted from the ~~simulations of the~~ 7000 SWAT model combinations were evaluated visually. We present ~~visualizations of the results of the GSA and the simulation uncertainty bands~~ approaches for the visualization of the simulation uncertainties that proved to ~~be powerful diagnostic tools in this study~~support the diagnosis of how the analyzed inputs affected the simulation of discharge and NO_3^--N loads.

Based on the GSA we identified climate change and the model parametrization to be the most ~~sensitive~~ influential model inputs for the simulation of discharge and NO_3^--N loads in both case studies. In contrast, the impact of the model setup on the simulation of discharge and NO_3^--N loads was low and the changes in land use and point source emissions were found to have the ~~least~~ lowest impact on the simulated discharge and NO_3^--N loads. ~~Additionally, the~~ The visual analysis of the uncertainty bands illustrated that the anomalies in precipitation of the different climate scenarios dominated the changes in simulation outputs, ~~rather than changes~~ while the differences in air temperature ~~in both case studies~~showed no considerable impact.

1 Introduction

Environmental systems are under constant change. Predicting the development of natural resources in a changing system involves large uncertainties (Milly et al., 2008). Climate change, in concurrence with other dynamic processes such as population growth, land use change or economic development pose challenges to the management of water supply and water quality (Duran-Encalada et al., 2017; Yates et al., 2015). Human disturbances can exacerbate the impacts of climate and amplify consequences to water quality (Jiménez et al., 2014) on one hand. On the other hand, stakeholders in environmental systems have to respond to future changes, for instance adapting farm management practices due to changes in temperatures and precipitation patterns (Schönhart et al., 2018). Ideally, an impact assessment considers all future changes that can affect the development of the environment of interest as well as those future changes that can introduce uncertainties in the simulation of the environmental variables of interest.

Changes in environmental systems are typically represented by discrete scenarios in impact studies. Preferably, the set of scenarios representing a dynamic change covers the full range of trajectories along which the development is plausible (Clark et al., 2016). Scenario development involves different data sources and models, which can introduce and propagate uncertainties. For example, climate scenarios have several sources of uncertainty and may include several socioeconomic scenarios (e.g. the current "Representative Concentration Pathways" (RCP; Moss et al., 2010; van Vuuren et al., 2011)) that drive an array of global climate models (GCM) (Knutti and Sedláček, 2013). However, the GCMs also have inherent uncertainty and they provide the boundary conditions for regional climate models (RCM) (e.g. Jacob et al., 2014). Further, the downscaling (Wilby et al., 1998; Wood et al., 2004) of the RCM simulations and the bias correction (Teutschbein and Seibert, 2013, 2012) are associated with their own uncertainty and are a standard procedures in climate scenario development. Eventually, it is essential to characterize the uncertainties inherent in all processes that affect the simulation of an environmental variable.

To simulate the development of hydrological variables under changing conditions, the developed scenarios are implemented as boundary conditions in hydrological models that are calibrated for historic observations. Yet, often different model setups and different sets of parameters in a model can perform equally well to reproduce historical observations of the variables of interest. Equifinality is a well-known issue in hydrologic modeling that has been extensively addressed in the literature (e.g. Schulz et al., 1999; Beven, 2006; Beven and Freer, 2001; Beven, 1996), where multiple model structures (e.g. Clark et al., 2008) and model parametrizations (e.g. Schulz et al., 1999) represent observations equally well and thus cannot be rejected (Beven, 2006). An adequate representation of historical data does not necessarily assure that different model setups agree when extrapolating to future conditions (Chiew and Vaze, 2015; Milly et al., 2008). Thus, differences in the model setup are a source of uncertainty in the simulation of an environmental variable under future conditions.

Altogether, an impact study comprises an abundance of combinations of trajectories of system changes and model setups to describe an environmental system that ultimately characterize the uncertainties in a simulation. Hence, a comprehensive description of the uncertainties in model simulations is a major challenge of any impact study. ~~An assessment is only as good as the dominant contributors of uncertainty in such a modeling chain.~~

Model sensitivity analysis (SA) can be used to derive the impact of different input variables on hydrological target variables. SA investigates the response of a modeled variable to the variation of model input variables (Saltelli et al., 2004). For a local sensitivity analysis (LSA) the model inputs are varied around a point (often an ~~“optimum”~~ ’optimum’ point) in the model input space. Global sensitivity analysis (GSA) assesses the sensitivity of a model output for the entire feasible range of model inputs (Gupta and Razavi, 2017; Pianosi et al., 2016). Compared to LSA, GSA usually requires a larger number of computations. Thus, a substantial part of recent GSA literature focuses on the computational efficiency and the robustness of GSA methods (e.g. Pianosi and Wagener, 2015; Razavi and Gupta, 2016a; Sarrazin et al., 2016; Cuntz et al., 2015; Rakovec et al., 2014), but also on increasing the insight into modeled systems from a certain number of model ~~computations~~ evaluations (e.g. Borgonovo et al., 2017; Dai et al., 2017; Guse et al., 2016a; Massmann et al., 2014; Razavi and Gupta, 2016a).

The complexity and computational demand of a model determine the feasible number of model evaluations and thereby the applicability of a SA method (Razavi and Gupta, 2015). Large atmospheric model applications, for instance, only allow a LSA with a few model evaluations (Gupta and Razavi, 2017; Pianosi et al., 2016). Environmental model applications are usually less computationally expensive and allow a more extensive GSA, illustrated in many environmental modeling studies (e.g. Guse et al., 2016b; Haghnegahdar et al., 2017; Massmann and Holzmann, 2015; Razavi and Gupta, 2016b; Sarrazin et al., 2016). Most applications utilize GSA to identify ~~and rank continuous model parameters~~ influential model parameters and to rank model parameters according to their influence on model outputs. Model parameters are usually continuous model inputs. (Saltelli et al., 2008; Baroni and Tarantola, 2014).

Although ~~,~~ it is possible to implement ~~more complex model inputs in~~ composite model inputs (e.g. climate scenarios that affect several climate variables at the same time, or land use scenarios that can impact the entire model setup) in a GSA and therefore to employ GSA in impact studies, a consideration of discrete and ~~complex~~ composite model inputs can constrain the applicability of GSA and complicate the implementation (Baroni and Tarantola, 2014). In impact studies, the response of an environmental variable to a (future) change in a model input is usually inferred by implementing a scenario realization of the respective model input in a model setup. From an SA perspective, this approach is equivalent to a local ~~“~~’one-at-a-time~~”~~’ (OAT) assessment of the model input sensitivity (Saltelli and Annoni, 2010; Baroni and Tarantola, 2014). ~~An~~ A local OAT analysis however presumes linear models and non-correlated inputs which are hardly true for any environmental model application (Rosolem et al., 2012; Baroni and Tarantola, 2014). Thus, to account for interactions of model inputs and model non-linearities the application of GSA is recommended instead (Saltelli and Annoni, 2010; Saltelli and Tarantola, 2002; Baroni and Tarantola, 2014).

Yet, a few studies implemented discrete and ~~complex~~ composite model inputs in GSA. With the Generalized Probabilistic Framework, Baroni and Tarantola (2014) rendered a solid basis for the implementation of correlated, non-continuous model inputs in GSA and applied the variance-based SA ~~(Sobol, 1993)~~ method of Sobol (1993) to assess the response of soil moisture, evapotranspiration, and soil water fluxes to uncertainties in meteorological input data, crop parameters, soil properties, model structure, and observation data. In a synthetic example, Dai and Ye (2015) performed model and scenario averaging to assess the impact of different model structures and scenarios of precipitation on groundwater flow and reactive transport in the soil. In a more recent study, Dai et al. (2017) employed ~~variance-based SA (Sobol, 1993)~~ the method of Sobol to identify the relevant system processes for groundwater flow and reactive transport represented in different model structures. Savage et al. (2016) applied GSA to identify the dominant controls in the calculation of flood inundation, to assess whether a high spatial resolution of the flood inundation model or the model parametrization is dominating the simulation. The mentioned studies illustrate the use of GSA with discrete and ~~complex~~ composite model inputs. ~~No study is known to us that takes advantage of GSA in the scope of environmental impact studies.~~ Anderson et al. (2014) and Butler et al. (2014) highlight the importance of assessing the uncertainty of future climate change impacts and the identification of relevant drivers and their interactions for climate policy making.

In this paper we demonstrate the utility of GSA ~~in~~ and uncertainty analysis in a comprehensive setting of an environmental model impact ~~studies~~ study and address the following points:

- We apply GSA ~~to~~ in two environmental modeling impact studies to identify the dominant sources of uncertainties for the simulation of discharge and nitrate-nitrogen (NO$_3^-$-N) loads~~using~~ . We analyze the impacts of different spatial aggregations of the model setup and different model parametrizations ~~, also while applying changes to~~ and assess the effects of changes in the land use, point source emissions, and the future climate.

- We ~~accompany the GSA with a visual analysis of the simulation uncertainties of~~ analyze the resulting uncertainties in the simulation of the long-term monthly mean discharge and monthly sums of NO$_3^-$-N loads, as well as flow duration curves (FDCs) of daily discharge and daily NO$_3^-$-N loads visually. We present ways to visualize the discrete model inputs that provide further insights into the relationships of uncertainties ~~between the model inputs and the simulated uncertainties~~in the simulations and different properties of the discrete realizations of the model inputs.

- Based on the GSA and the visual analysis of the ~~simulation~~ simulated uncertainties we are able to draw conclusion ~~for the implemented model setups, model parametrizations~~ on the simulation of discharge and NO$_3^-$-N loads as impacted by the model setup, model parametrization and the future scenarios of land use, point source emissions and ~~the climateconcerning their impact on the simulation of discharge and (NO$_3^-$-N) loads and on the assumptions we~~ climate. These conclusions are of course limited to assumptions made in the ~~description of the uncertainties in the~~ model setup and in the development of the scenarios.

The paper is structured in the following way: Section 2 contains an overview of the two investigated catchments, the Soil and Water Assessment Tool (SWAT, Arnold et al., 1998) that we implemented in this study, and the preparation of the model input data that we used in the model setup. In Section 2.4 we describe the setup of the SWAT model with different spatial aggregations and illustrate the pre-processing of the SWAT model setups that was necessary to identify the sensitive SWAT model parameters and to define non-unique parameter sets for all model setups. The scenarios of land use, point source emissions and the climate together with the input data and pre-processing to develop the individual scenarios are specified in Section 2.5. Section 2.6 combines the SWAT model setups, the defined non-unique model parametrizations and the developed scenarios of land use, point source emissions and climate in the GSA and explains the methods we applied to analyze the sources of uncertainties for the simulation of discharge and NO_3^--N loads. The results of the combined GSA framework and the visual analysis are provided in Section 3. We discuss the findings of the GSA application and the visual analysis of the simulation uncertainties for the two case studies in Section 4 and address the specific assumptions that we made during the model setup and the development of the scenarios.

2 Materials and Methods

2.1 Study sites

The two investigated catchments (Schwechat and Raab) are representative examples for river systems for the eastern region of Austria. Both rivers have their origin in the forested foothills of the limestone Alps with a pre-alpine character and a low anthropogenic impact. The lower parts of both catchments are characterized by human activities, with primarily urban settlements and agricultural uses in the plains of the Schwechat catchment and dominant industrial activities and agricultural land uses in the valley bottom of the Raab catchment (Fig. 1 and Tables ~~A1 and A2~~A3 and A4).

The Schwechat river has its source in the Vienna woods at the northeastern boundary of the Northern Limestone Alps with a maximum altitude of 893 m a.s.l. After a natural flow section in the narrow and dominantly forested valley of the “Helenental” (70% of the total catchment area. See Table ~~A1~~A3), the Schwechat drains into the Vienna basin with flat topography and a predominance of agriculture, viniculture and settlement areas. The main agricultural crops are winter wheat and summer wheat. Larger areas in the upper part of the catchment are used as pastures (~10% of the total area). The largest settlement is the city of Baden with a population of approximately 26000 inhabitants, while smaller settlements are scattered over the catchment. All municipal wastewaters are collected in three wastewater treatment plants (WWTP, black triangles in Fig. 1), where the the WWTP Baden is the most relevant one with a capacity of 45000 population equivalents (PE). All WWTPs perform carbon removal, nitrification, denitrification and enhanced phosphorus removal. Due to the close proximity to the city of Vienna population growth is a likely prospect for the settlement areas in the lower part of the catchment. The part of the catchment considered in this study has its outlet next to the city of Traiskirchen at an altitude of 185 m a.s.l. and covers an area of approximately 275 km². The long term mean annual precipitation in the Vienna Basin is around 620 mm/yr and the mean annual temperature is 9.9°C.

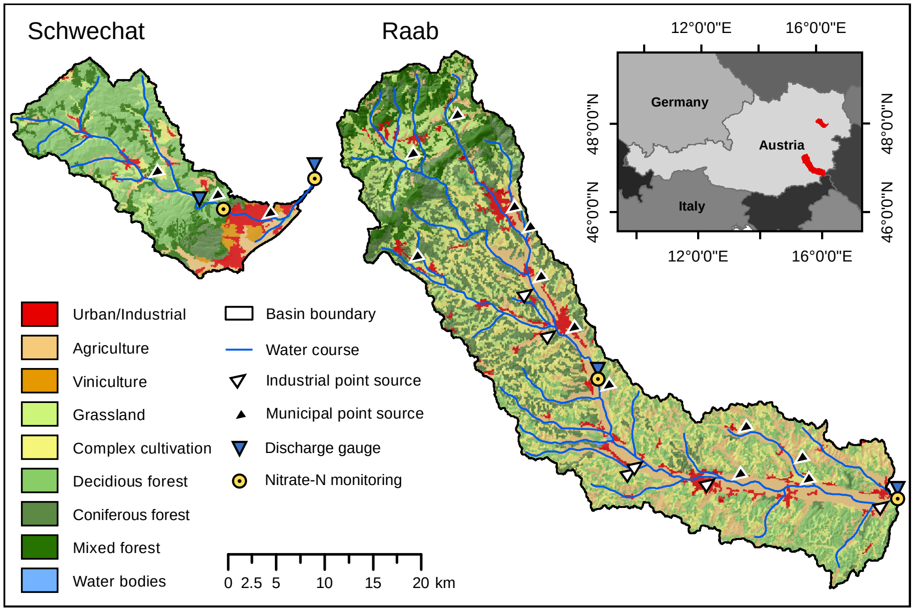

Figure 1. Study sites Schwechat (left) and Raab (right).

The Raab river originates at the edge of the southeastern Alps. These are characterized by low mountain ranges with a maximum altitude of 1547 m a.s.l., mostly covered by forests (~42% of the total catchment area. See Table ~~A2~~A4). The Raab flows through the southern part of Austria and crosses the boarder to Hungary close to the city of Neumarkt an der Raab at an altitude of 232 m a.s.l. The case study encompasses the Austrian part of the Raab with a catchment area of approximately 998 km². The long-stretched river valley is dominated by agricultural activities (~25 % of the total area), with urban areas in between. The slopes along the Raab are covered with heterogeneous patterns of forests, pasture areas and agricultural land use. The main agricultural crops are corn and oil seed pumpkins, but also wheat and vegetable production are common. While the urban areas are of similar small structure as in the Schwechat catchment, leather industries are present in the catchment that release substantial nutrient inputs into the receiving waters, which has resulted in ~~transboundary~~ trans-boundary conflicts (Ruzicka et al., 2009). Municipal wastewaters in the Raab catchment are collected in 12 relevant WWTPs (black triangles in Fig. 1) that all have the same standards for ~~wasterwater~~ wastewater treatment as in the Schwechat catchment, but have almost three times the total capacity (approximately 150000 PE). Six relevant industrial emitters are located along the main reach of the Raab river (white triangles in Fig. 1) that all perform internal waste water treatment following the respective industry-specific regulations for wastewater treatment (e.g., BGBl. II Nr. 10/1999, 1999; BGBl. II Nr. 12/1999, 1999). The average annual precipitation in the Raab catchment is approximately 800 mm/yr and the long term annual mean temperature is 9.0°C.

2.2 The Soil and Water Assessment Tool (SWAT)

The SWAT model (Arnold et al., 1998) is a continuous, process based, semi-distributed eco-hydrological model. In this study we implemented SWAT2012 (Rev.622) to simulate daily time series of discharge and NO_3^--N loads at the catchment outlets. The models' spatial reference to a catchment is given by a subdivision of the basin into subbasins. Areas containing the same land use, soil type and lying in the same slope range are lumped together in each subbasin to form hydrologic response units (HRUs). All processes on the land phase of each subbasin are calculated at the HRU scale and are further propagated into the water phase of each subbasin. The processes calculated on the land phase include water balance components such as interception, infiltration, shallow and deep percolation, surface runoff, lateral flow, groundwater flow, plant uptake and evapotranspiration, or the pathways of nutrients such as the input through atmospheric deposition, or ~~fertilzier~~ fertilizer application, the transformation into other forms of a nutrient and the transport ~~thruough~~ through surface runoff, percolation, lateral flow and return flow in the groundwater (Neitsch et al., 2011). In the water phase, the nutrients budgets are calculated. Following the calculation of the water balance and the nutrient budgets, the discharge, the nutrient loads and other substances are routed through the linked subbasins to the defined catchment outlet (Neitsch et al., 2011). The required input data to set up a model with SWAT are a digital elevation model (DEM), a raster land use map including the model parametrization and the performed ~~managemement~~ management operations for each land use, a raster soil map with soil physical and chemical parameters for all soil layers, and meteorological input data.

2.3 Model input data and data preparation

A DEM with a 10 m resolution was available for Austria from an airborne laser scan (Geoland.at, 2015). Based on the DEM we defined three slope classes with slopes of 0-3%, 3-8%, and >8% in the HRU definition step.

CORINE land cover (EEA, 2015) served as the base land use map to which more detailed agricultural data was added. CORINE does not classify agricultural land uses into crop types. Therefore, tabular data of agricultural land uses at the municipal level derived from the 2010 Austrian agronomic census (Statistik Austria, 2015b) was superimposed onto CORINE data by randomly distributing crops according to the crops' areal share at the municipal level to CORINE pixels containing agricultural and complex cultivation land use. Typical time windows for planting, fertilizer application, tillage and harvest were derived from field experiment records for the individual crops (Land NÖ, 2015) and written to the HRU management files. The management dates were randomized for all HRUs within the time windows derived for a management operation. Dates with strong rainfall or a high soil moisture potential were not used for scheduling management operations. With 70.0% and 42.3% forest land uses where the most dominant land uses in the Schwechat and the Raab catchments, respectively. The SWAT model setups differentiated between deciduous forests, coniferous forests and mixed forests, derived from CORINE land cover (see Tables ~~A1 and A2~~A3 and A4). All HRUs with one of the three forest types as land use were parameterized with an initial biomass and an initial leave area index to simulate intact forests in both catchments.

The SoilGrids data base (Hengl et al., 2017) is a consistent global soil information system that provides soil physical and chemical parameters at a 250m grid resolution and seven soil depths. We utilized the available soil parameters from SoilGrids and estimated further required soil parameters with pedo-transferfunctions provided by the R package euptf (Tóth et al., 2015). The seven available soil depths from the SoilGrids data were aggregated to three soil depths (0-30cm, 30-100cm, and 100-200cm), and the gridded data were clustered into soil classes applying kmeans clustering (R Core Team, 2017, Hartigan and Wong (1979)) resulting in 14 and 8 "optimum" soil classes for the rivers Schwechat and Raab respectively.

Meteorological input data was available from the INCA system developed and operated by the Central Institute for Meteorology and Geodynamics of Austria (ZAMG; Haiden et al., 2011). INCA provides reanalysis data of precipitation and temperature on 1km grid resolution for Austria with a temporal resolution of 15 minutes for precipitation and 60 minutes for temperature in the period from 2003 to 2015. For all SWAT model setups, daily precipitation sums and daily minimum and maximum temperatures were temporally and spatially aggregated for the model subbasins.

Point source emission data was available from external emission monitoring of municipal WWTP greater than 2000 PE according to BGBl. 1996/210 (1996) for both catchments. Municipal WWTP larger than 2000 PE are responsible for 99.2% and 86.3% of municipal point source emissions in the Schwechat and the Raab catchments respectively. Thus, these data cover a substantial part of the municipal emissions. Additionally, daily and weekly internal monitoring data was available for some large WWTP schemes. In most cases however, only information on NO_3^--N emissions was provided. A general budgeting of nitrogen emissions however showed, that the substantial share of total nitrogen is emitted in form of NO_3^--N(87% in the Schwechat catchment and 89% in the Raab catchment). For industrial emitters monthly and annual records from internal and external monitoring agencies were available and only allowed an estimation of industrial emissions with coarse temporal resolution, while covering the annual budgets. Again, mainly data for NO_3^--N emissions were available. Although, nitrogen is emitted in different forms the available ~~databasis~~ data basis only allowed to consider NO_3^--N loads contributed by point sources.

Table 1 provides an overview of the model input data that was used for the SWAT model setup.

Hourly observations of discharge were available for the period from 2003 to 2015 at two gauges for the Schwechat and the Raab each (Fig. 1). NO_3^--N concentration readings with varying time intervals of 5 to 15 minutes were available at two stations in both catchments (yellow circles in Fig. 1) for selected time periods resulting from monitoring campaigns at the rivers Schwechat (BMLFUW, 2013) and Raab (BMLFUW, 2015a, b). SWAT simulates output variables with daily time steps. To compare the observations with the modeled SWAT outputs of discharge and NO_3^--N loads, daily NO_3^--N loads and daily mean discharge were calculated from the observation data.

2.4 Model setup, parameter selection and identification of non-unique parameter sets

~~The study takes into account the effect that the SWAT model setup and model parameterization has on the simulation of discharge and NO_3^--N loads.The SWAT model setups for the Raab and the Schwechat involved decisions for the selected~~ Graphical GIS user interfaces such as ArcSWAT (Winchell et al., 2015) or QSWAT (Dile et al., 2016) facilitate the setup of SWAT models. Yet, a model setup requires the modeler to define the number of subbasins as well as the number of HRUs (e.g. by removing HRUs with areas below a certain threshold from the setup and apportion their areas to the remaining HRUs). The size and the number of subbasins ~~of~~ in a model setup ~~and the definition of the HRUs. Both modifications are necessary~~

Table 1. Input data for the SWAT model setup, the data sources, and data processing steps.

Input data set	Data source	Data preparation
Topography	DEM Austria (Geoland.at, 2015)	Digital Elevation Model for Austria in 10m resolution.
Land use	CORINE Landcover (EEA, 2015), 2010 Austrian agronomic census (Statistik Austria, 2015b)	Basis: CORINE Land cover, Agricultural areas ~~resampled~~ re-sampled with statistical information from 2010 Austrian agronomic census.
Soil data	soilgrids.org (Hengl et al., 2017), euptf (Tóth et al., 2015)	Basis: SoilGrids 250m resolution in 7 depths. Clustered in space and and aggregated over depth. Further SWAT soil parameters derived using pedotransfer functions.
Meteorology	INCA (Haiden et al., 2011)	Preciptation and temperature data in 1km resolution.
Agricultural practices	Statistik Austria (2015b), Land NÖ (2015)	Derive time periods and sequences of field management practices from field experiments.
Point source emissions	External monitoring, Internal records of WWTPs	Time series and point measurements of discharge and NO_3^--N concentrations.

~~decisions for any SWAT model setup. For both catchments we identified the relevant SWAT model parameters employing GSA and selected parameter sets that adequately reproduced historical observations of discharge and NO_3^--N loads.~~ can affect the process simulations and the resulting model outputs (Jha et al., 2004; Momm et al., 2017; Tripathi et al., 2006). Removing small HRUs from the model setup and allocating their areas to the remaining HRUs affects the distribution of land use, soil types, and slope classes and thus can impact the model simulations substantially (Jha et al., 2004).

~~For the SWAT model setup we~~ We used the ArcSWAT plugin (Version2012.10_1.14~~; (Winchell et al., 2015)~~) together with ArcGIS 10.1 (ESRI, 2012) ~~involving DEM, land use, soil and meteorological data~~for the model setup. For both case studies we set up the SWAT model with ~~a different number~~ different numbers of subbasins, whereby we prepared model setups with the full number of HRUs and respective setups with a reduced number of HRUs for each catchment. ~~The size and the number of subbasins in a model setup can affect the process simulations and the resulting model outputs (Jha et al., 2004; Momm et al., 2017; Tripathi~~ ~~. Therefore, eliminating HRUs that have an area below a certain threshold and allocating their areas to the remaining HRUs will affect the distribution of land use, soil types, and slope classes and thus can affect model simulations substantially (Jha et al., 2004), yet, it is common practice in setting up a SWAT model.~~

In total, we set up four SWAT models, two with 3 and two with 14 subbasins for the Schwechat catchment and six ~~setups~~ models for the Raab catchments with two each of 4, 29, and 54 subbasins. ~~To maintain a comparable aggregation error with the number of subbasins for the different model setups , we employed~~ For the full HRU setups we kept the resulting HRUs unmodified. For the model setups with a reduced number of HRUs we eliminated small HRUs. We determined thresholds for land use, soil, and slope classes to remove HRUs that have an area below these found thresholds. The thresholds

Table 2. SWAT model setups for the Schwechat and the Raab catchment including the numbers of subbasins and the number of HRUs for each setup.

	Schwechat			Raab	
Setup	# Subbasin	# HRU	Setup	# Subbasin	# HRU
sw_14_full	14	1434	rb_54_full	54	5349
sw_14_thru	14	196	rb_54_thru	54	954
sw_03_full	3	606	rb_30_full	30	3516
sw_03_thru	3	64	rb_30_thru	30	584
			rb_04_full	4	755
			rb_04_thru	4	115

were determined using the R package ~~topHRU (Strauch et al., 2016) to define of the thresholds for land use, soil,~~ ’topHRU’ (Strauch et al., 2016). ’topHRU’ enables to find thresholds that minimize the number of HRUs of a SWAT model setup while minimizing the aggregation error (sum of changes in the areas of land uses, soils and slope classes ~~and accepted a maximum~~ of the reduced set of HRUs compared to the full HRU setup). To maintain a comparability between the reduced HRU setups thresholds were selected that result in an aggregation error of 5% ~~of the total area of the HRUs. Table 2~~ in all reduced HRU model setups. Table 2 gives an overview of the final ~~baseline~~ model setups for both case studies.

In a ~~pre-analysis step~~parameter screening, we applied a GSA to the simulations of discharge and NO_3^--N loads at the catchment outlets of all SWAT model setups ~~individually to identify the relevant~~ to identify influential model parameters. ~~Starting with the same~~ Initially, 42 ~~parameters, we~~ model parameters were selected that are frequently calibrated in SWAT model setups to simulate discharge and NO_3^--N loads. The SWAT model setup initializes the model parameters using values obtained from the SWAT data bases (either standard values or user defined, e.g. by pedotransfer functions). The selected initial ranges to modify the model parameters and the selected types of parameter changes (e.g. replace parameter values globally or modify a spatially distributed parameter field by a fraction of a parameter) reflect typical procedures often found in SWAT model calibration studies. An overview of the model parameters that were identified as influential and that were further used in the model impact study is provided in Table A1.

We employed the STAR VARS approach (Razavi and Gupta, 2016a, b) ~~using the IVARS50 measure~~ to screen and rank the model parameters. STAR VARS utilizes variograms along each model input dimension of the input space to infer each model inputs influence on a target variable over different scales (where short lag distances approximate the derivative based method of Morris (Morris, 1991) and long distances the method of Sobol (Sobol, 1993)). The calculation of the variograms is based on the tailored STAR sampling design where "star center" points are randomly sampled in the input space. For each center point cross sections are sampled along the input factor dimensions with an equally spaced interval. For each sampled input combination the model is evaluated and variograms along the response surface are calculated. Razavi and Gupta (2016a) proposed integrated measures of the variograms as measures of sensitivity, where the measures $IVARS_{10}$, $IVARS_{30}$, and $IVARS_{50}$ represent the integrals over 10%, 30%, and 50% of each input dimension respectively and therefore provide the sensitivity of a target variable to a model input over different scales. A detailed description of the method is provided in Razavi and Gupta (2016a) and the STAR sampling is outlined in Razavi and Gupta (2016b). The method proved to be robust and computationally efficient for high dimensional problems (e.g., Razavi and Gupta, 2016b; Haghnegahdar et al., 2017; Sheikholeslami et al., 2019; Haghnegahdar and Razavi, .

We drew STAR samples (Razavi and Gupta, 2016b) with 50 center points and ten parameter samples per parameter dimension ~~were drawn resulting~~ that resulted in 18950 parameter combinations per model setup. ~~We employed~~ To calculate the target variables we used the Nash Sutcliffe Efficiency criterion (NSE, Nash and Sutcliffe, 1970), the Kling Gupta Efficiency criterion (KGE), including its three components (Gupta et al., 2009), and a refined version of the Index of Agreement (Willmott et al., 2012) to evaluate the ~~daily~~ simulated time series of ~~discharge and~~ daily mean discharge and daily sums NO_3^--N loads~~, and~~ . Additionally, we applied the ratio of the root mean square error and standard deviation (RSR, (Moriasi et al., 2007)) to evaluate different segments of the FDCs of daily discharge and daily NO_3^--N load simulations (Pfannerstill et al., 2014; Haas et al., 2016). A model parameter was considered to be sensitive if it showed a relative sensitivity of 10% compared to the most sensitive parameter with respect to a specific objective criterion for at least one of the employed objective criteria.

The ~~GSA conducted for the parameters identified the same parameters to be sensitive parameters for all model setups for the Schwechat~~ performed GSA for the model parameters of the different model setups of the Schwechat catchment and the Raab ~~catchments, respectively ,~~ catchment respectively showed very similar results independent of the number of subbasins and HRUs of the individual model setups (Fig. A1). Therefore, for the impact study the same set of model parameters was considered as influential for all model setups of the Schwechat and the Raab, respectively. In total, 19 ~~and 16 sensitive parameters were identified~~ parameters for the Schwechat and 16 parameters for the Raab ~~, respectively (Table 3~~were identified to be influential for the analyzed target variables (Table A1).

~~Sensitive SWAT model parameters for the model setups of the Schwechat and the Raab. Parameter Description Schwechat Raab SFTMP Snowfall temperature (°C) X X SNOCOVMX Minimum snow water content that corresponds to 100% snow cover. X X SNO50COV Snow water equivalent that corresponds to 50% snow cover X SURLAG Surface runoff lag time (h) X X GW_DELAY Groundwater delay (d) X X GW_REVAP Groundwater revaporation coefficient X X GWQMN Treshold depth of water in the shallow aquifer required for return flow to occur (mm) X X RCHRG_DP Deep aquifer percolation fraction X X SOL_K Saturated hydraulic conductivity (mm/h) X X SOL_AWC Available water capacity of the soil layer X X SLSOIL Slope length for lateral subsurface flow X X CANMX Maximum canopy storage X ESCO Soil evaporation compensation factor X X LAT_TTIME Lateral flow travel time X X OV_N Manning's n-value for overland flow X CNOP_till SCS runoff curve number for the tillage operation X X RCN Concentration of nitrogen in rainfall X X NPERCO Nitrogen percolation coefficient X X CDN Denitrification exponential rate coefficient X SDNCO Denitrification threshold water content X~~

~~For all SWAT model setups of the~~ To represent the model parametrization as an input in the subsequent sensitivity and uncertainty analysis of the environmental impact study, non-unique parameter sets were identified for the Schwechat and the Raab catchments, respectively. The preceding parameter SA revealed that changes in the model parameter values influenced the simulations similarly independent of the subbasin and HRU configurations in the Schwechat and the Raab ~~catchments we~~

~~identified non-unique parameter sets that adequately simulated daily observation of~~ catchment, respectively. As a consequence, but also to facilitate the separation of the effects of the model setup and the model parametrization in the analysis, we selected parameter combinations as non-unique ones that result in simulations of daily discharge and NO_3^--N loads ~~. From the sensitive model parameters of each case study we drew~~ that fulfill certain objective criteria together with all model setups of the Schwechat and the Raab, respectively. For the respective 19 and 16 influential model parameters we randomly sampled 100000 parameter ~~sets using a generic random sampling and applied the 100000 parameter sets to all SWAT model setups~~ combinations and simulated daily discharge and NO_3^--N loads with all model setups of the Schwechat and the Raab catchments. We evaluated the simulations with the following criteria to accept a parameter set: KGE > 0.5 for daily discharge at the catchment outlets, KGE > 0.4 for daily NO_3^--N loads at the gauges with longer continuous records (in both case studies the gauging point within the catchment and not at the catchment outlet), percentage bias (Gupta et al., 1999) < 50% for NO_3^--N loads, and absolute RSR < 1 for different discharge and NO_3^--N load (according to Pfannerstill et al., 2014; Haas et al., 2016). In total, we identified 43 and 52 behavioral parameter ~~sets~~ combinations for the Schwechat and the Raab catchments, respectively. The ability of the selected parameter sets ~~together~~ used with the different model setups to reproduce the observed data is illustrated in Fig. A2. The initial and final ranges of parameter changes are shown in Table A2. The 43 and 52 parameter combinations are additionally illustrated in parallel coordinate plots for the Schwechat and the Raab in Fig. A3 to show any clustering of individual parameters and interactions between parameters.

2.5 Scenario definition

The study involves future changes of the land use, point source emissions, and the climate. The uncertainties of these variables are expressed as discrete scenarios.

For the land use change scenarios, two scenario ~~storylines~~ story lines (Rounsevell and Metzger, 2010) were developed for the Schwechat and the Raab catchments. A "business-as-usual" scenario extrapolates the observable trends in land use change to the future (2071 to 2100), while a second "extensive" scenario assumes an extensification of agricultural activities and other intensive land uses in both catchments (Table ~~A3~~A5).

In the Schwechat catchment population growth is the strongest factor for a future change in land use (Statistik Austria, 2015a, 2016). Hence, a transformation from extensive pasture land (-35%) to urban land use and an increase of dense urban areas describe the "business-as-usual" scenario. The "extensive" scenario assumes no change in population and a shift of half of the wheat producing area to extensive pastures.

Since 1970, the areas for corn production increased by 220% in the Raab catchment, mostly for biogas production and at the expense of sugar beets and cereals (Statistik Austria, 2017). For the "business-as-usual" scenario, an increase in the corn area by a further 100% until the end of the century was assumed, replacing extensive pastures (-75%), sugar beets (-80%), legumes (-70%), and winter wheat (-30%).

Groundwater protection measures lead to strict regulations for fertilizer application in the Leibnitzerfeld region adjacent to the Raab catchment (LGBl. Nr. 39/2015, 2015). Therefore, the "extensive" scenario assumes an adoption of similar nitrogen regulations in the Raab catchment. Thus, decreasing areas with intensive fertilizer application, such as corn by 50% and transforming these areas to extensive pasture land was carried out in this scenario.

Two municipal point source emission scenarios for both case studies (Table ~~A4~~A6) and two industrial point source emission scenarios for the Raab catchment (Table ~~A5~~A7) were developed. The future change in municipal emissions was assumed to be directly related to the change in population. For all provinces in the Schwechat basin future scenarios predict an average population growth of 32% (Statistik Austria, 2015a, 2016). The predictions of the population development in the provinces of the Raab are contradicting, with predicted changes between +2.3% (Statistik Austria, 2015a) and -20.4% (Amt d. Stmk LReg, 2016).

In the Raab catchment 94% of the industrial point source emissions stem from the leather industry and almost 70% of the industrial point source emissions are caused by one leather manufacturing company. Thus, industrial emission scenarios were developed for that particular manufacturer. As boundaries for the production, we defined an upper environmental boundary and a lower economical boundary for the prediction of future industrial emissions. Based on an assessment of effluent dilution (ÖWAV, 2010), current environmental regulations (BGBl. II 2010/99, 2010; and BGBl. II 2006/96, 2006) allow an increase of 30% in emissions from that leather producer, resulting in a total increase in industrial emissions of 22.6%. Assuming a relocation of the two manufacturing sites of that leather producer to outside of the catchment would stop their emissions into the Raab, reducing the total industrial point emissions by 75.2%.

Future climate change was considered with 22 downscaled and bias corrected climate change scenarios (Table ~~A6~~A8). Regional climate simulations were obtained from the EU-CORDEX project (Jacob et al., 2014), providing 11 GCM-RCM simulations for the emission scenarios RCP4.5 (Smith and Wigley, 2006; Wise et al., 2009) and RCP8.5 (Riahi et al., 2007) respectively. In this study we utilized daily precipitation sums and daily minimum and maximum temperatures for the time period 2071 to 2100. The EURO-CORDEX climate simulations are available at a spatial resolution of 12.5 km (EUR-11) (Jacob et al., 2014). Statistical downscaling (Zorita and Von Storch, 1999) was applied to prepare all climate simulations at a resolution of 1 km. To correct downscaling errors (e.g. Haslinger et al., 2013; Muerth et al., 2013), bias correction (Teutschbein and Seibert, 2013) was applied to the climate simulations employing quantile mapping (Hempel et al., 2013). Downscaling and bias correction were performed for the historical period 1971 to 2000, involving the reanalysis datasets SPARTACUS (Hiebl and Frei, 2016) for minimum, mean and maximum temperature and GPARD (Hofstätter et al., 2013) for daily precipitation sums.

2.6 Analysis

Table ~~??~~ 3 summarizes the land use change, point source emissions, and climate change and the model setups and model parametrizations that were used for the analysis of simulated discharge and NO_3^--N loads in the Schwechat and the Raab catchments. In total, 7000 combinations of land use, point source emissions, climate, model setups and model parametrizations were drawn for both case studies applying a ~~generic~~ quasi random sampling. The number of combinations results from previous experiments that applying the SA method of Sobol (results not shown) using the sampling strategy proposed by Saltelli and Tarantola (2002) with a base sample size $N_b = 1000$ and a total sample size of $N = N_b(k+2)$, where k is the

Table 3. SWAT inputs implemented in the sensitivity analysis case studies and their numbers of discrete realizations for the Schwechat and the Raab catchments.

Input	# Values		Details on values
	Schwechat	Raab	
Land use scenario	2	2	one "extensive", one "business-as-usual"
Point source scenario	2	4	Population growth: optimistic/pessimistic , Industry Raab: production increase/resettlement
Climate scenario	22	22	11 RCP4.5, 11 RCP8.5, period: 2071-2100
Model setup	4	6	Raab: 54, 30, 4 subbasins with/without HRU reduction, Schwechat: 14, 3 subbasins with/without HRU reduction
Parametrization	43	52	KGE discharge >0.5, KGE NO_3^--N >0.4, pbias NO_3^--N <50%

number of model inputs that are analyzed. All sampled combinations were assembled to executable SWAT models. Daily discharge and daily NO_3^--N loads at the outlets of the Schwechat and the Raab catchments were simulated for the period from 2071 to 2100.

The analysis of discharge and NO_3^--N loads follows two main goals i) to identify the dominant controls on the simulation of discharge and NO_3^--N loads in the two case studies and ii) to assess how the considered inputs control the simulation of discharge and NO_3^--N loads.

2.6.1 Global sensitivity analysis

To ~~identify the impact~~ measure the relative importance of the developed model input scenarios, the model setup and the parametrization on the simulation of daily discharge and daily NO_3^--N loads, we employed GSA using the PAWN sensitivity index (Pianosi and Wagener, 2015). PAWN ~~involves~~ employs the empirical cumulative distribution function (CDF) of a target variable to infer the model input ~~sensitivity (Pianosi and Wagener, 2015). Thus, PAWN is applicable to discrete model inputs. Further,~~ influence (Pianosi and Wagener, 2015). PAWN is moment-independent and was found to be a robust measure for sensitivity of non-symmetrically distributed outputs of environmental models (Pianosi and Wagener, 2015; Zadeh et al., 2017).

PAWN expresses the sensitivity of a target variable x to a model input x ~~for a target variable y~~ by computing a distance measure between the unconditional CDF $F_y(y)$ (where all model inputs are perturbed) and the conditional CDF $\cancel{F_{(y|x_i)}(y)}$ $F_{(y|x_i)}(y)$ (where the model input of interest is fixed and all others are perturbed). ~~The distance measure~~ Pianosi and Wagener (2015) proposed is the Kolmogorov-Smirnov test statistics as a distance measure. The distance $\cancel{KS_j(x_i)}$ $KS_j(x_i^j)$ between the CDFs for the model input x_i fixed at a value $x_i = x_i^j$ is defined as:

$$KS_j({x_i}^j) = \left\| F_y(y) - F_{y|x_i, x_i = x_i^j}(y) \right\|_y \quad (1)$$

To assess the overall sensitivity considering all fixed values of x_i, the values of ~~$KS_j(x_i)$~~ $KS_j(x_i^j)$ are summarized for all j sampling points. A summary statistics (Pianosi and Wagener (2015) suggested e.g. median or maximum) is applied to compute the PAWN index T_i for the model input x_i. ~~Due to the characteristics of the influencing variables~~ The model inputs that are analyzed in this study ~~(large differences in the number of values for each input) employing the maximum statistics is advantageous, as it provides an understanding of a maximum possible sensitivity induced by an input variable , rather than providing information on~~ strongly differ in their numbers of discrete realizations. Further, the distribution of the ~~average sensitivity caused~~resulting Kolmogorov Smirnov distances can be highly skewed (e.g. the majority of discrete realizations has a low impact, while a few realizations strongly influence the simulation). Therefore, the significance of an average sensitivity of a target variable y_i to a model input x_i is questionable. In a setting where the strongest impact of a model input x_i on a target variable y_i is of major interest the application of a maximum statistics is beneficial. Hence, the PAWN sensitivity index is defined here as:

$$T_i = \max_{\cancel{x_i=x_i^1 \dots x_j^{n_i}}\; x_i=x_i^1 \dots x_i^{n_i}} (KS_j({x_i}^j)) \tag{2}$$

The ~~discrete values $x_i, x_i^1, \dots, x_i^j, \dots, x_i^{n_i}$~~ values $x_i = x_i^1, \dots, x_i^j, \dots, x_i^{n_i}$ are the n_i ~~possible states~~ discrete realizations of the input x_i. The resulting PAWN sensitivity index varies between 0 and 1 where a lower value of T_i implies a lower ~~sensitivity~~ influence of the input x_i on the target variable y.

~~The sampling scheme of PAWN (Pianosi and Wagener, 2015) was initially designed~~ Pianosi and Wagener (2015) introduced the PAWN sensitivity method using a specifically tailored sampling design to infer the PAWN indices T_i for continuous model ~~parameters, which requires a modification for discrete model inputs . The initial sampling~~ inputs x_i. The proposed sampling scheme suggests to draw N_c conditional samples at ~~n~~ n randomly sampled points of each influencing variable x_i, where x_i is fixed at a value $x_i = x_i^j$ while all others are perturbed. ~~Instead,~~ Recently, Pianosi and Wagener (2018) extended the applicability of the PAWN sensitivity method to estimate T_i from a generic random sample of continuous model inputs. To approximate T_i the generic sample N is split into n segments along each model input dimension resulting in conditional samples N_c with an approximate size of N/n. We employed the proposed updated sampling strategy and adapted it for the use with discrete model inputs. A sample of the size N was drawn~~and subsetted with N/n_i subsets for all~~. For each model input combination every model input was sampled randomly from its discrete realizations. To infer $KS_j(x_i)$ for all discrete values x_i^j ~~to infer the values for all $KS_j(x_i)$~~of a model input x_i the sample N was split into subsets for all n_i discrete values, resulting in subsets of the size N/n_i on average. It is important to consider, that the subset size depends on the number of discrete values n_i of a model input x_i, while the subsets resulting from the sampling scheme proposed by Pianosi and Wagener (2018) had an average size of N/n for all model inputs x_i.

To account for the effect of different numbers of discrete realizations of the analyzed inputs, but also to assess whether the number of drawn samples of input combinations ($N = 7000$) was sufficient to perform a GSA with PAWN, confidence intervals were calculated for the PAWN indices applying bootstrapping (Hinkley, 1988; Efron, 1987) using the R package

’boot’ (Canty and Ripley, 2017). To calculate the bootstrap mean and the 95% confidence intervals, 1000 bootstrap replicates were drawn (as demonstrated in Sarrazin et al. (2016)).

Signature measures of discharge and NO_3^--N loads were used ~~for the sensitivity assessment~~as target variables y. Signature measures are measures that describe specific characteristics of simulated time series (Euser et al., 2013) (in this case of daily mean discharge and daily sums of NO_3^--N loads). We calculated quantile values (0.01, 0.05, 0.20, 0.70, 0.95, and 0.99) of daily discharge and daily NO_3^--N loads, long-term mean discharges and long-term mean sums of NO_3^--N loads on an annual basis and for the meteorological seasons spring, summer, autumn, and winter, and mean NO_3^--N concentrations for different ranges of discharge quantiles (very high discharge (above 0.95 quantile), high discharge (0.95 to 0.70 quantile), medium discharge (0.70 to 0.20 quantile), low discharge (0.20 to 0.05 quantile), and very low discharge (below 0.05 quantile)).

2.6.2 Visual analysis of the simulation uncertainties

To investigate how the inputs of land use change, changes in point source emissions, climate change, the model setup or the model parametrization control the simulation of discharge and NO_3^--N loads, we analyzed the simulation outputs and their associated uncertainties visually. The 7000 assembled combinations of model inputs, model setups and parametrizations resulted in ranges of simulated discharge and NO_3^--N loads. ~~In this study, we consider all~~ All executed model setups ~~to be plausible~~ represent plausible realizations of the future conditions in both catchments to simulate future discharge and NO_3^--N loads. Thus, the overall simulation uncertainties of simulated discharge and NO_3^--N loads comprise all 7000 simulations of the Schwechat and the Raab catchments, respectively.

We visually analyzed the uncertainty bands (no thresholds were set) of the simulations of the long-term mean monthly specific discharge, the long-term mean monthly sums of NO_3^--N loads and the FDCs of daily discharge and daily NO_3^--N loads. These variables are related to a wide range of the signature measures that were analyzed in the GSA and thus allow a comparison of the GSA results with the results of the visual uncertainty analysis.

The low number of ~~each input included in the study~~ possible values taken by each input allowed a more detailed analysis of their effect on the simulated uncertainties, by subsetting the uncertainty bands of the discharge and NO_3^--N load simulations with respect to the individual realizations of the analyzed model input. The separated simulation uncertainty bands were additionally colored with respect to the specific properties of an input, such as the temperature or precipitation anomalies of each climate scenario compared to historical records. These color ranges greatly facilitated identifying the dominant controls of the simulation.

3 Results

3.1 Sensitivity analysis

~~PAWN indices were calculated for~~ Fig. 2 summarizes the influence of the implemented land use, point source emission, climate scenarios, the different model setups and ~~model parametrizations employing the calculated signature measures (Section 2.6.1)~~

~~of the simulated daily times series of~~ the model parametrizations for the simulation of future discharge and NO_3^--N loads ~~.~~ ~~Together, the~~ in the Schwechat (left) and the Raab (right) catchments. Each plot panel shows the calculated PAWN indices for ~~one of the analyzed inputs provide~~ the analyzed target variables for one model input in a catchment. Related target variables are grouped by colors to support the interpretability (e.g. to identify changes in sensitivity from high to low discharge). In its entity each panel provides a general overview of the importance of ~~the respective~~ an input for the simulation of discharge and NO_3^--N loads. Individual PAWN indices (a single bar in a plot panel) highlight the importance of an input for the simulation of specific characteristics of the time series of discharge and NO_3^--N loads~~(Fig. 2)~~.

The white boxes on top of each bar show the bootstrap means and the 95% confidence intervals (CI) of each PAWN index and therefore provides an indicator for the adequacy of the sample size that was used to perform the analysis and the impact of differing numbers of discrete values of the analyzed input variables. In general the bootstrapping resulted in narrow confidence intervals (maximum +0.05 and -0.08) for all analyzed model inputs and all signature measures providing high confidence in the resulting sensitivities. Although the numbers of discrete realizations of the analyzed model inputs (e.g. only 2 land use scenarios, but 43 and 52 model parametrizations) differ strongly and therefore result in different subset sizes to calculate the PAWN indices, no substantial differences in the confidence intervals is visible.

The land use scenarios applied to SWAT demonstrated a rather negligible impact on all signature measures, with mean PAWN indices below 0.05 and 0.07 and confidence intervals in the same range for the Schwechat and Raab respectively (first row Fig. 2). The point source scenarios, in contrast, showed a considerable influence on the signature measures of NO_3^--N loads and concentrations in the Raab case study, while the ~~sensitivities~~ impacts of the point sources in the Schwechat case study were negligibly low (second row Fig. 2). Thus, based on the implemented point source emission scenarios, industrial emitters in the Raab catchment are relevant for the development of in-stream NO_3^--N loads and concentrations, particularly for low discharges and low NO_3^--N loads. The importance of the industrial point sources in SWAT increases when higher NO_3^--N load quantiles (low NO_3^--N loads, from dark yellow to light yellow in Fig. 2)) and NO_3^--N concentrations for low discharges (from dark red to light red in Fig. 2) are simulated, which is evident from an increase in the mean PAWN index from 0.11 to ~~0.5~~ 0.49 and 0.22 to ~~0.40~~0.43, respectively. The climate scenarios and the model ~~setups~~ parametrizations show respective decreases in their importance for the simulation of low NO_3^--N loads and NO_3^--N concentrations for low discharges ~~,~~ (with decreases in the mean PAWN index from ~~0.72 to 0.30 and~~ 0.71 to 0.28 for the climate scenarios' influence on NO_3^--N loads and from 0.79 to ~~0.42~~0.36 for model parametrization's influence on NO_3^--N concentrations).

The implemented climate scenarios showed large ~~sensitivities for~~ impacts on all calculated signature measures of discharge and NO_3^--N loads (third row Fig. 2). The mean PAWN indices range between ~~0.31 to 0.92 and 0.29~~ 0.25 to 0.90 and 0.25 to 0.96 for the Schwechat and the Raab, respectively. The climate scenarios were the most relevant inputs for the simulation of seasonal mean discharges and seasonal sums of NO_3^--N loads. For the simulation of low discharge quantiles (large daily discharges) climate scenarios showed the highest relevance. For the simulation of low discharges however, the importance of the climate scenarios decreases, while the model parametrization becomes more relevant (from dark green to light green in Fig. 2). The mean PAWN indices of climate scenarios drop from ~~0.78~~ 0.74 to 0.47 in the Schwechat catchment and from ~~0.89~~

~~to 0.56~~ 0.82 to 0.51 for the simulation of lower discharges, while the mean PAWN indices for the model parametrization show respective increases from ~~0.54~~ 0.43 to 0.87 and ~~0.53 to 0.83~~0.44 to 0.80.

In general, the model parametrization was highly ~~sensitive~~ influential for all calculated signature measures and is comparable to that of the climate scenarios, with mean PAWN indices ranging between 0.43 to 0.90 in the Schwechat and ~~0.29 to 0.96 and 0.43 to 0.83~~ 0.36 to 0.80 in the Raab (fifth row Fig. 2). Particularly, for the simulation of NO_3^--N concentrations the model parametrization was the most dominant control of the variable simulated. In contrast to the large ~~sensitivities~~ impact of the model parametrization, the relevance of the model setup was much lower for the simulation of discharge and NO_3^--N loads and concentrations. Overall, values of the PAWN index for the choice of the model setup did not exceed ~~0.39~~0.37, and were much smaller (two to five times) compared to the model parametrization. ~~For most of the analyzed signature measures the~~ The model setups yielded insignificantly low PAWN indices for the majority of signature measures with values below 0.1 in the Raab case study (2.5 % CI almost 0 for many signature measures), indicating that the model setup ~~was not sensitive. Although the Raab case study shows low sensitivities for the model setups, their overall sensitivities follow the general trend of the climate scenarios to a large extent. In particular, for large~~ had a low influence on most of the analyzed processes. Only for high discharges and large NO_3^--N loads ~~the model setup shows a higher sensitivity, indicating a relationship of climate driven runoff and NO_3^--N transport in SWAT~~a mean value for the PAWN index above 0.1 is visible.

3.2 Analysis of the simulation uncertainties of discharge and NO_3^--N loads

Using all 7000 combinations of land use, point source emissions, climate, model setups, and model parametrizations, the simulated discharges and NO_3^--N loads deviated by up to 350% (grey bands in Fig. 3) from the simulations of discharge and NO_3^--N loads in the reference period 2003 to 2015 (dashed line in Fig. 3). In the Schwechat (left column in Fig. 3) ~~as well as the Raab case study~~ wider uncertainty bands are visible for the spring and early summer months. The results for the Raab catchment (right column) ~~wider uncertainties~~ show wider uncertainty bands emerged for ~~the summer months~~ summer as well as for winter/early spring. A ~~substantial difference that is visible for~~ notable difference between the two case studies is how the ~~reference period relates to the uncertainty bands in amplitude~~simulations of long term monthly discharges and NO_3^--N loads in the reference period compare to the ranges of future simulations. While the majority of model combinations for the Schwechat simulated larger discharges and NO_3^--N loads for all months in the future, for the Raab catchment the simulations of discharge and especially NO_3^--N loads are lower in comparison to the reference period.

The ~~analysis~~ analyses of the uncertainty bands with respect to the implemented land use scenarios and the point source scenarios fully ~~confirms~~ confirm the results from the SA (Fig. 4). The attributed uncertainty bands for the two land use scenarios almost entirely overlap and show only minor deviations. A similar result is illustrated for the two point source scenarios in the Schwechat case study. The scenarios in the Raab catchment ~~that~~ involved industrial point source emissions~~however, show a difference of approximately 15 tons per month in~~ . The grouped uncertainty bands that include scenarios with an increase in industrial production (red) and the uncertainty bands that include a decrease in industrial production (blue) show similar patterns. Yet, the blue and red uncertainty bands show a clear shift to each other. On average the scenarios with an increase in industrial production show long-term monthly sums of NO_3^--N loads ~~, this is due to an increase in production, or a production~~

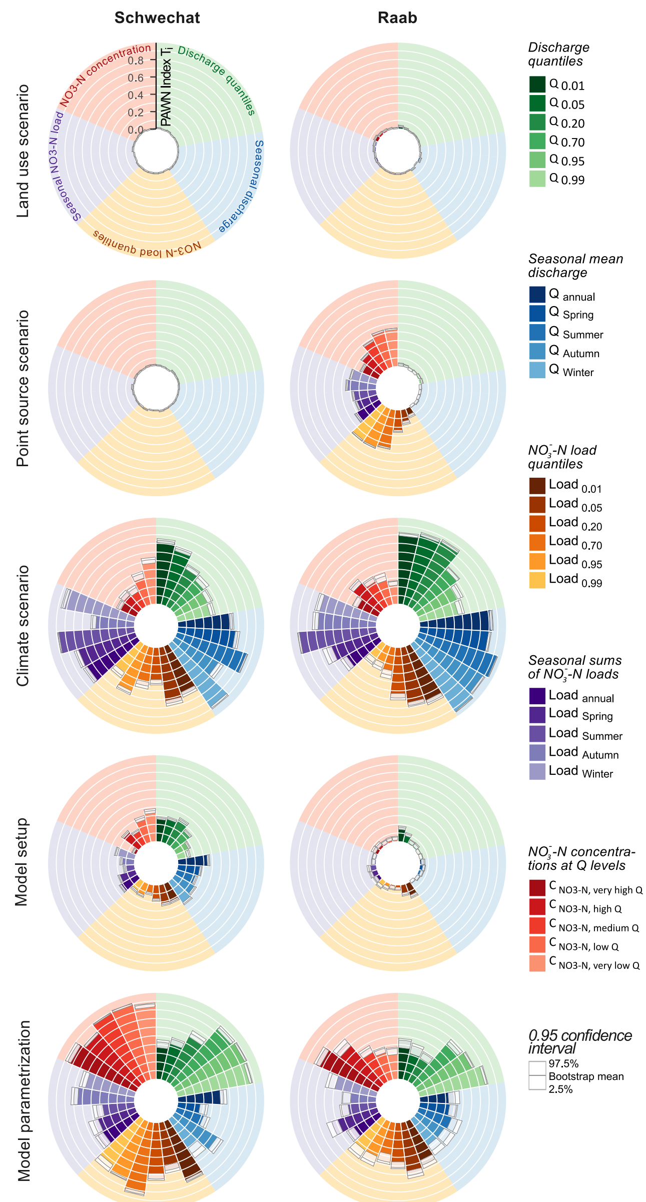

Figure 2. ~~Model input sensitivities for~~ Sensitivities of signature measures of discharge and $NO_3^- - N$ loads in the Schwechat (left) and the Raab (right) catchment ~~. For each row~~ to the ~~sensitivities of~~ model inputs land use scenarios, point source scenarios, climate scenarios, the model setup, and the model parametrization~~are plotted~~. Each circle plot shows the set of PAWN indices calculated for the respective case study and model inputs. ~~The~~ PAWN indices are illustrated in colored groups ~~showing in~~ and clockwise order ~~the sensitivities of selected~~ for discharge quantiles ~~in~~ (green), ~~of~~ seasonal long-term mean discharges ~~in~~ (blue), ~~selected~~ quantiles of $NO_3^- - N$ loads ~~in~~ (yellow), seasonal sums of $NO_3^- - N$ loads ~~in~~ (purple), and ~~the~~ mean $NO_3^- - N$ concentrations for discharge quantiles ~~in~~ (red). The white boxes represent the bootstrap mean and the 95% confidence intervals for the calculated PAWN indices.

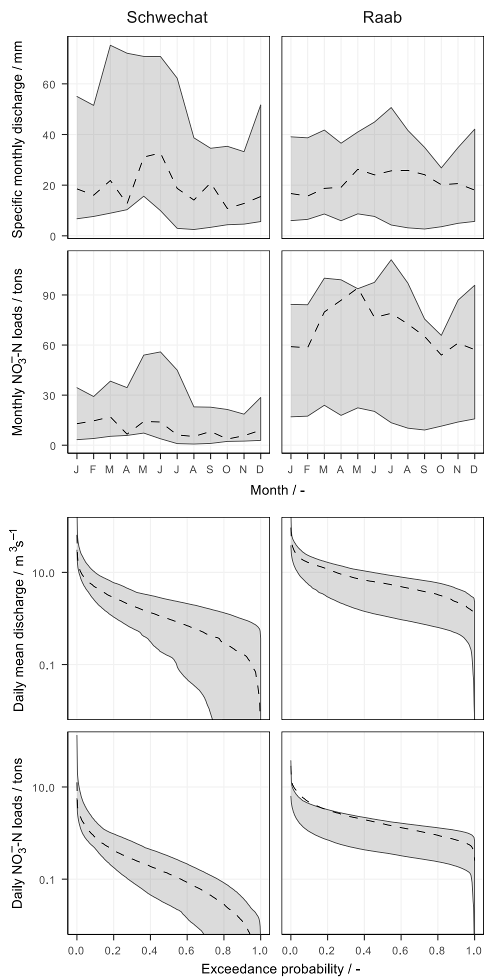

Figure 3. Simulated uncertainties resulting from the 7000 combinations of realizations of the influencing variables for the Schwechat (left) and the Raab (right). The grey bands illustrate the absolute ranges of simulated long-term mean monthly specific discharge (first row), long-term monthly sums of $NO_3^- - N$ loads (second row), FDCs of mean daily discharges (third row), and FDCs for daily sums of $NO_3^- - N$ loads (fourth row). The dashed lines show the best simulation of the historical reference period.

~~stop of the major leather producer in the region~~that are 15 tons higher compared to the scenarios with a decrease in industrial production. The same scenarios show larger amplitudes for medium and low NO_3^--N loads, while large NO_3^--N loads remain uninfluenced by the two scenarios for the development of the leather industry.

~~The GSA~~ With the GSA we identified the climate scenarios to have a great influence on ~~the sensitivity~~ all signature measures of the simulated variables~~in all simulations~~. Attributing the uncertainty bands to the individual GCM-RCM combinations unveils diverse outcomes for the future flow regime, the distribution and amplitude of monthly NO_3^--N loads, as well as the appearance of high and low discharges and NO_3^--N loads (Fig. 5). A visual analysis of the separated uncertainty bands identifies the mean annual precipitation anomalies of the GCM-RCM combinations to have a strong impact on the simulation of discharge and NO_3^--N loads. In comparison to the reference period (dashed line), wetter future climate scenarios (blue) simulated larger discharge and NO_3^--N loads, while dryer future conditions lead to a drastic reduction in discharge and NO_3^--N loads.

Half of the 22 implemented GCM-RCM combinations simulated an increase of more than 75 mm (dark blue) and for two GCM-RCM combinations, an increase of more than 25 mm (light blue) of precipitation for the Schwechat catchment was simulated. In contrast, for the Raab nine and four GCM-RCM combinations simulated a decrease in precipitation of more than 75 mm (dark red) and 25 mm (light red), respectively. Consequently, a decrease in discharge and NO_3^--N loads due to a decrease in precipitation is pronounced in the Raab catchment, while the majority of simulations of the Schwechat catchment show an increase in discharge and NO_3^--N loads.

While a grouping of the individual climate scenarios with respect to their temperature anomalies shows a more indefinite picture~~. All~~ , all climate scenarios simulated an increase in temperature. Nevertheless, the expectation that an increase in annual mean temperature increases evapotranspiration and thus reduces discharge and NO_3^--N loads is not met in Fig. 6. A clear separation of warmer and cooler climate scenarios, as it is observable for precipitation is not the case with temperature. Consequently, the differences in precipitation predominantly account for the ~~sensitivities~~ influence of the climate scenarios, rather than the differences in temperature.

Although the influence of the model setups was much lower compared to the ~~sensitivities~~ influence of the climate scenarios or the model parametrization, the analysis of the uncertainty bands for the different model setups provides interesting insights (Fig. 7). The uncertainty bands do overlap to a great extent, which confirms a low impact of the use of different model setups in the simulation of discharge and NO_3^--N loads. Noteworthy is, that model setups that use the full set of HRUs agree much stronger in their simulations compared to the model setups where the number of HRUs was reduced. The difference between the full HRU and the reduced HRU model setups is distinct in the Schwechat case study. The uncertainty bands of the two full HRU model setups almost completely overlap, although their numbers of subbasins are different (4 and 14 subbasins). The two model setups with a reduced number of HRUs (but also with 4 and 14 subbasins) show differences of up to 15 mm in the simulated monthly specific discharge and up to 7 tons in the monthly NO_3^--N loads (~20 % of the uncertainty bandwidth).

The model parametrizations were relevant for all signature measures of discharge and NO_3^--N loads and were most dominant for medium and low flows. The most dominant model parameters in both case studies were the parameters CNOP_till and SOL_AWC. Both parameters control the water retention and thus the immanent contribution of rainfall to the river discharge. Large values of CNOP_till and small values of SOL_AWC reduce the water retention capacity and increase the amplitude

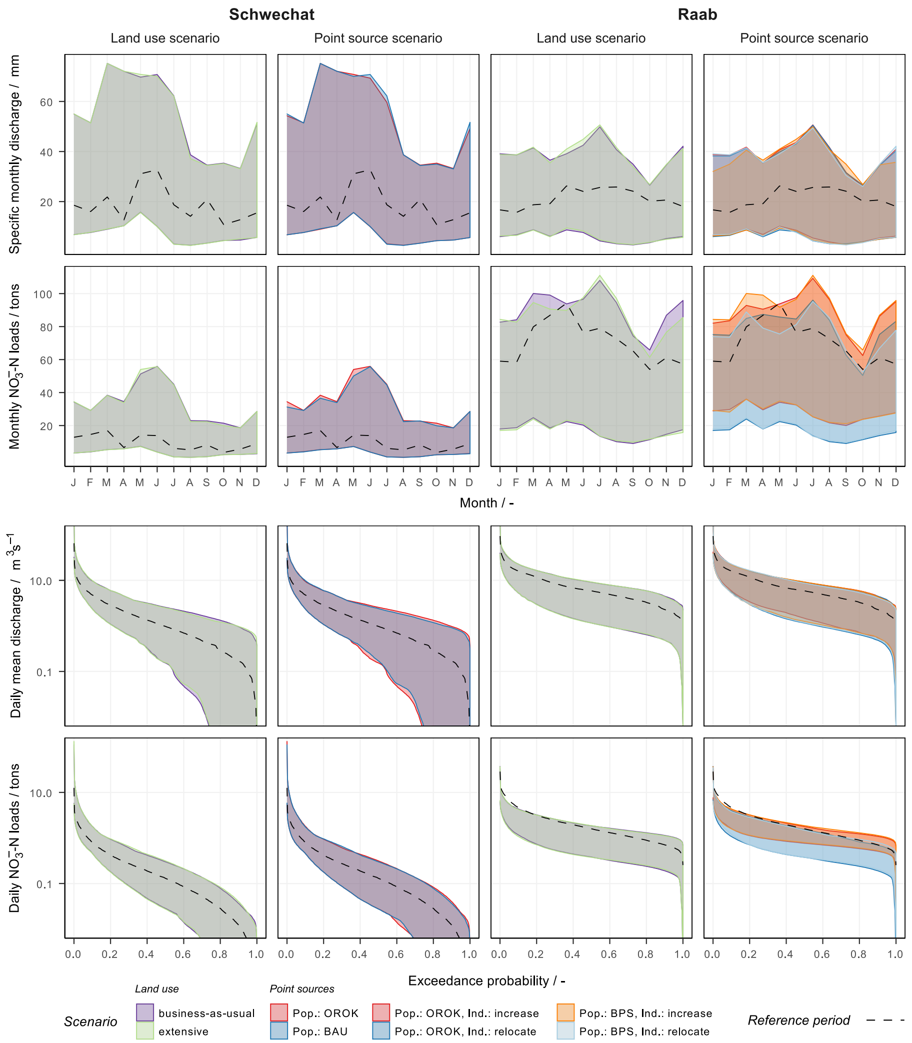

Figure 4. The influence of land use change and the development of point source emissions on the uncertainties resulting from the 7000 combinations of realizations of the influencing variables for the Schwechat (left) and the Raab (right). The uncertainties are illustrated for simulated long-term mean monthly specific discharge (first row), long-term monthly sums of $NO_3^- - N$ loads (second row), FDCs of mean daily discharges (third row), and FDCs for daily sums of $NO_3^- - N$ loads (fourth row). The uncertainty bands are attributed to the implemented land use scenarios (left panels per case study) and the point emission scenarios (right panels). The colors of the ~~subsetted~~ grouped uncertainty bands indicate the different scenarios. The dashed lines show the best simulation of the historical reference period. The corresponding land use changes are provided in Table A5. The corresponding population growth scenarios (Pop. in the legend) are listed in Table A6 and the corresponding industrial emission scenarios in the Raab catchment (Ind. in the legend) are listed in Table A7.

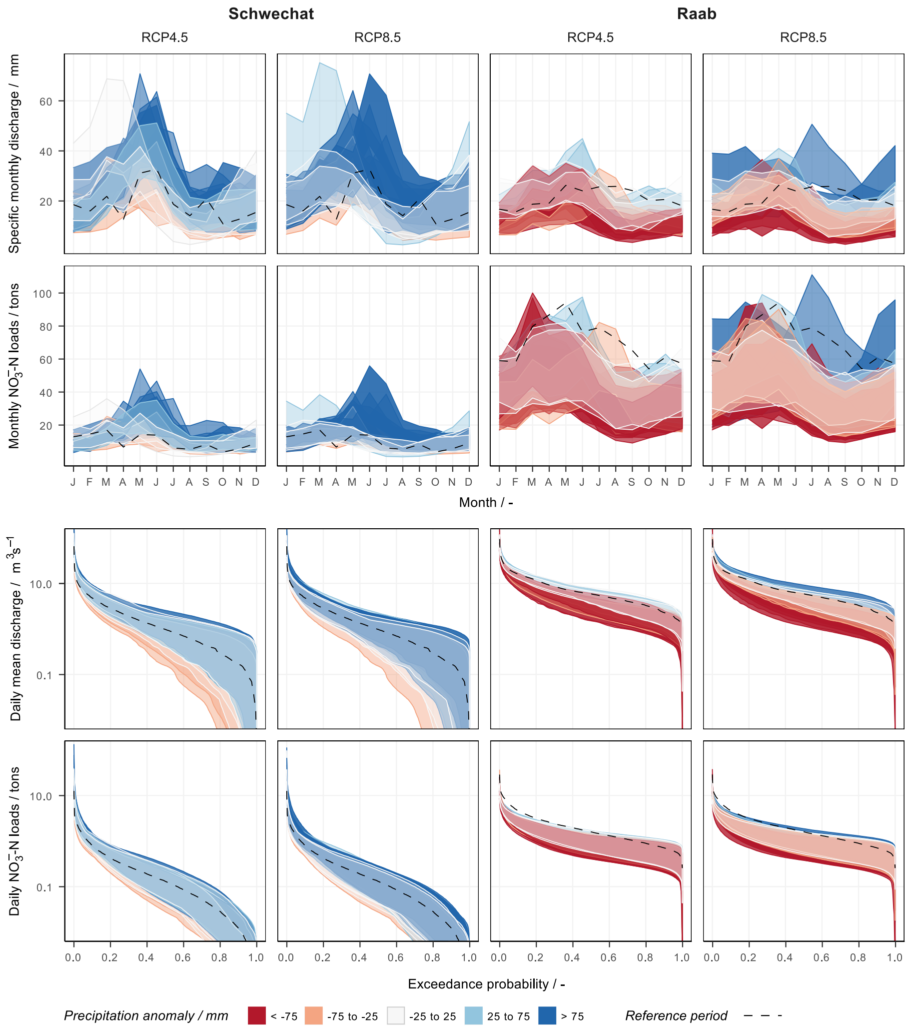

Figure 5. The influence of anomalies in precipitation on the uncertainties resulting from the 7000 combinations of realizations of the influencing variables for the Schwechat (left) and the Raab (right). The uncertainties are illustrated for simulated long-term mean monthly specific discharge (first row), long-term monthly sums of $NO_3^- - N$ loads (second row), FDCs of mean daily discharges (third row), and FDCs for daily sums of $NO_3^- - N$ loads (fourth row). The uncertainty bands are attributed to the individual implemented climate scenarios. The colors of the uncertainty bands show the anomalies in long-term mean annual precipitation of each climate scenario, where blue represents wetter conditions compared to the reference period and red dryer conditions. The dashed lines show the best simulation of the historical reference period.

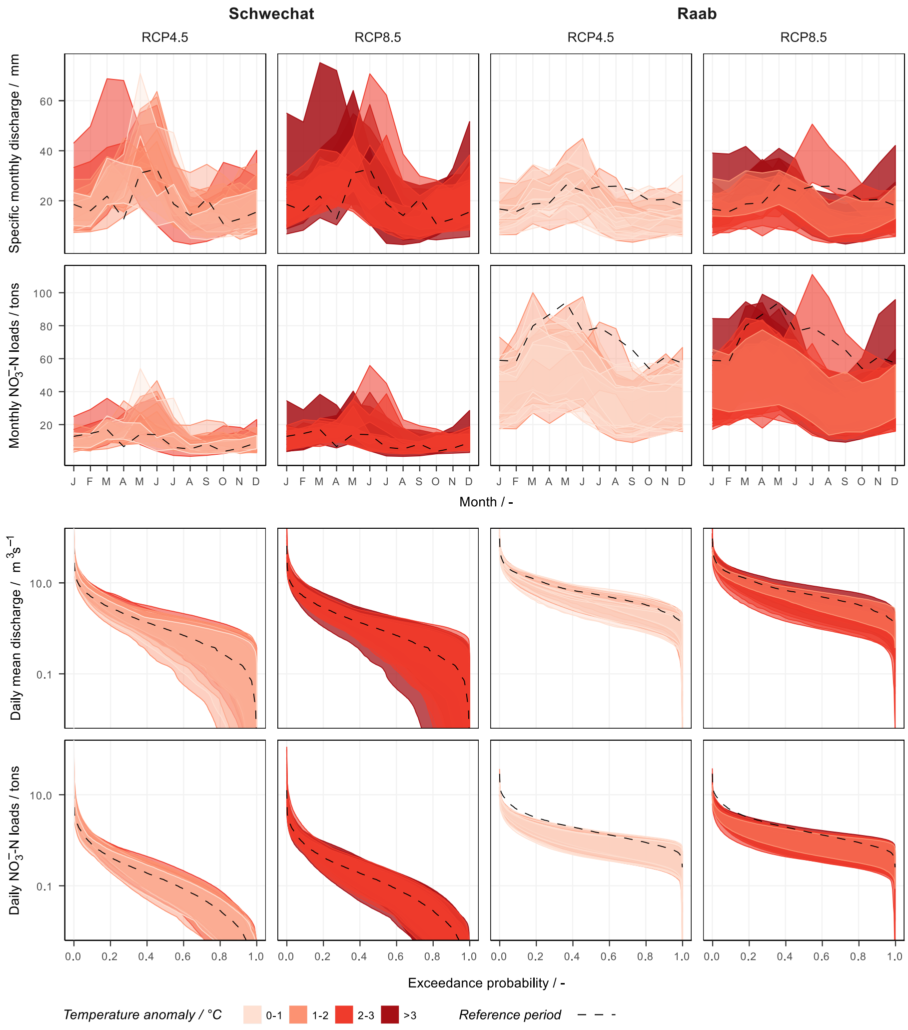

Figure 6. The influence of anomalies in air temperature on the uncertainties resulting from the 7000 combinations of realizations of the influencing variables for the Schwechat (left) and the Raab (right). The uncertainties are illustrated for simulated long-term mean monthly specific discharge (first row), long-term monthly sums of $NO_3^- - N$ loads (second row), FDCs of mean daily discharges (third row), and FDCs for daily sums of $NO_3^- - N$ loads (fourth row). The uncertainty bands are attributed to the individual implemented climate scenarios. The colors of the uncertainty bands show the anomalies in long-term mean annual air temperature of each climate scenario, where a darker red represents hotter conditions compared to the reference period. The dashed lines show the best simulation of the historical reference period.

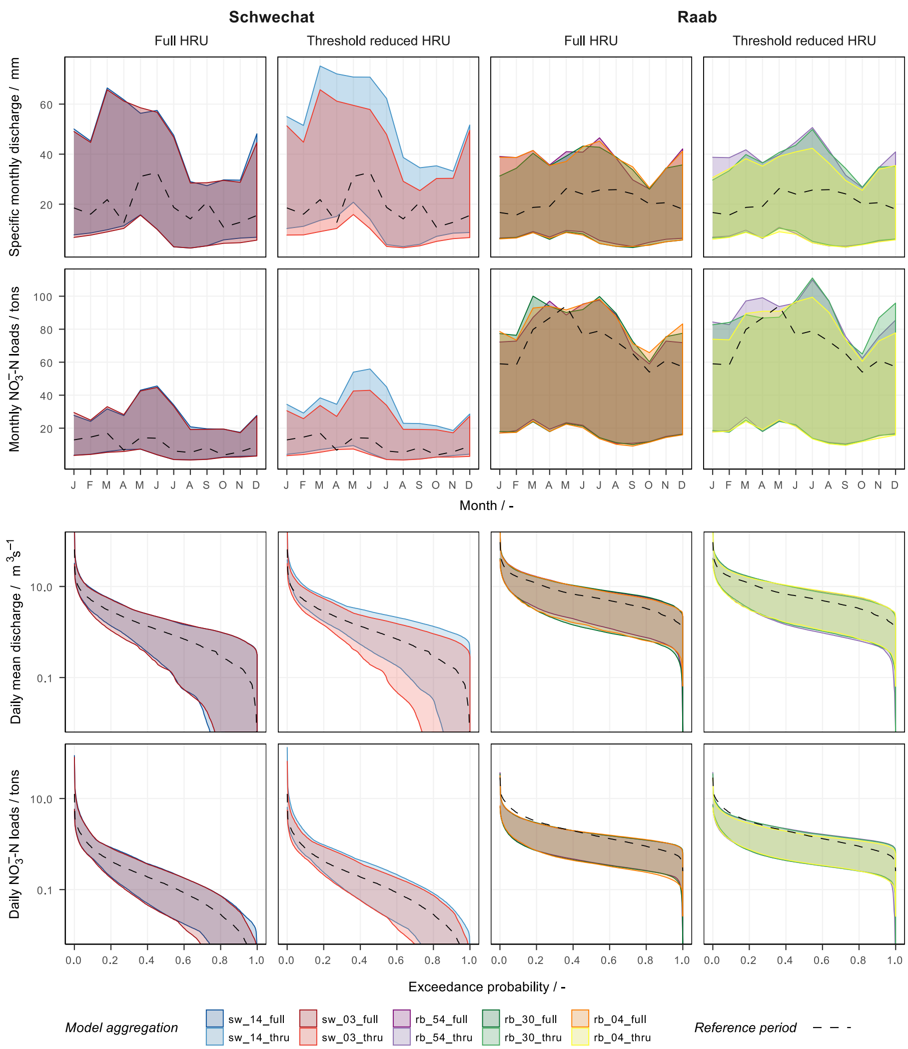

Figure 7. The influence of model setup on the uncertainties resulting from the 7000 combinations of realizations of the influencing variables for the Schwechat (left) and the Raab (right). The uncertainties are illustrated for simulated long-term mean monthly specific discharge (first row), long-term monthly sums of $NO_3^- - N$ loads (second row), FDCs of mean daily discharges (third row), and FDCs for daily sums of $NO_3^- - N$ loads (fourth row). The uncertainty bands are attributed to the individual SWAT model setups. The results are separated for model setups where the full set of HRUs was used (left panels per case study) and for setups with a reduced set of HRUs (right panels). The colors of the uncertainty bands show the different model setups with varying numbers of subbasins. The dashed lines show the best simulation of the historical reference period.

of medium and low discharges (third row in Fig. 8). A similar but inverse behavior is visible with medium NO_3^--N loads (last row in Fig. 8), where a higher water retention results in an increase of NO_3^--N loads. For the long-term monthly mean discharges and sums of NO_3^--N loads two effects are observable in Fig. 8. First, smaller values of CNOP_till and larger values of SOL_AWC decrease the upper boundary of the uncertainty bands. Second, selected model parametrizations with large values of CNOP_till and small values of SOL_AWC cause considerably larger discharges in spring and a strongly reduced runoff in the autumn months in the Schwechat case study.

4 Discussion

4.1 What can we as modelers learn from such analysis

The illustrated case studies emphasized the necessity to characterize, identify and explicitly communicate the uncertainties in a modeling chain, particularly for future simulations of environmental variables where large uncertainties are inherent in several modeling inputs. While the sensitivity analysis of signature measures related to discharge, NO_3^--N loads and NO_3^--N concentrations provided a comprehensive overview of the dominant influencing inputs on specific modeled variables, the analysis of the uncertainty bands for the simulation of the modeled variables provided insights into which properties of the model inputs (e.g. mean annual precipitation or mean air temperature of a climate scenario) control the uncertainties and how these control the simulation. The analyses allow to draw conclusions that are beneficial to consecutive steps of an impact study, for instance to refine the impact study setup and to focus on the most ~~sensitive~~ influential components and ultimately to reduce the uncertainties in the modeling simulation chain.

The land use scenarios showed an almost negligible impact on the simulation of discharge and NO_3^--N loads. The discharge and the NO_3^--N loads at the catchment are however integrated signals for the entire catchment and changes in land use may have a greater importance for particular points in a catchment. Many case studies have applied the SWAT model to assess the impact of land use change on different variables of the water cycle (Wagner et al., 2017; Mehdi et al., 2015b), water quality (Guse et al., 2015; Mehdi et al., 2015a), or sediment yield (Bieger et al., 2013). Bieger et al. (2013) found very low land use change induced increases in discharge for a catchment in China. Only an assumed strong intensification of the agriculture led to a 4% increase in discharge. At the same time however, a strong increase in sediment yield of up to 450% for the summer months was simulated due to the intensification of agriculture. Guse et al. (2015) also found only small changes in simulated discharge caused by future land use change in a German lowland catchment. In absolute numbers the simulated future NO_3^--N loads showed small differences between the baseline scenario and the two applied methods of land use change presented by Guse et al. (2015). Yet, the temporal patterns in NO_3^--N loads caused by the different approaches of changing the land use were the major observable difference. Mehdi et al. (2015b) in contrast, found that including future land use change into the impact assessment of a southern German watershed strongly increased the NO_3^--N and total phosphorus loads. In comparison, the low impact of land use change found in the present study seems reasonable, particularly as no extreme scenarios were implemented. Nevertheless, an assessment of whether the implemented scenarios adequately reflect the possible futures (e.g. fertilizer management) is recommended.

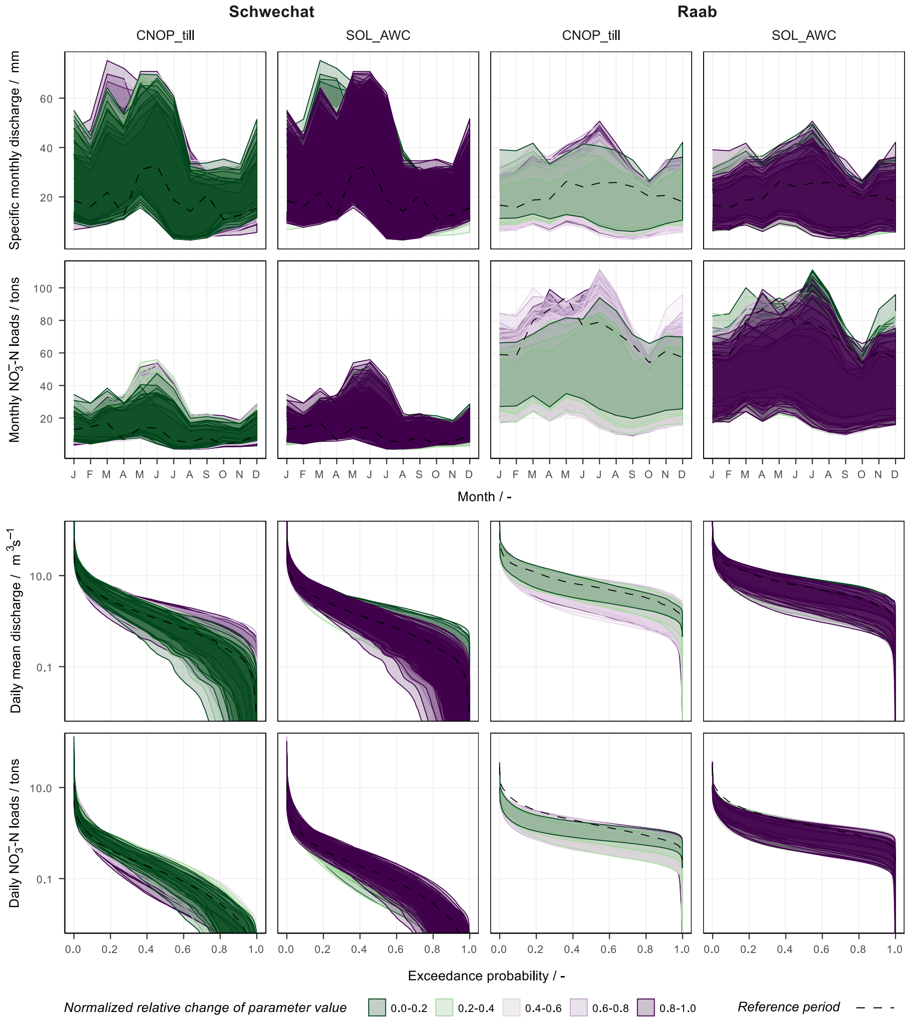

Figure 8. The influence of model parametrization on the uncertainties resulting from the 7000 combinations of realizations of the influencing variables for the Schwechat (left) and the Raab (right). The uncertainties are illustrated for simulated long-term mean monthly specific discharge (first row), long-term monthly sums of $NO_3^- - N$ loads (second row), FDCs of mean daily discharges (third row), and FDCs for daily sums of $NO_3^- - N$ loads (fourth row). The uncertainty bands are attributed to the individual 'behavioral' SWAT model parameter sets. The effect of the two dominant model parameters CNOP_till (left panels for each case study) and SOL_AWC (right panels) is shown. The subsetted uncertainty bands are colored with respect to the changes of the parameter values, shown as normalized values for comparability. The dashed lines show the best simulation of the historical reference period.

Industrial emitters were the main cause for the impact of point sources on medium to low NO_3^--N loads. The future scenarios of the development of industrial emitters were however highly uncertain. The developed scenarios are based on expert knowledge. Yet, there is no reliable basis available on status of the industrial emitters by the end of the century. Therefore, the developed scenarios should be noted as feasible futures, rather than e.g. politically realizable futures (Godet and Roubelat, 1996). To set a feasible range as boundaries for the future development of industrial emitters can lead to an overestimation of their ~~sensitivity~~ impact in comparison to other influencing variables. Nevertheless, the visualization of the NO_3^--N FDC of the Raab case study highlights the effect of the industrial emissions for medium and small NO_3^--N loads. Large NO_3^--N loads however, are hardly affected by the implemented scenarios, indicating that large NO_3^--N emissions are mainly driven by agricultural activities.

~~Climate~~ The selection of climate scenarios had a strong influence on the simulation of discharge and NO_3^--N loads in both case studies. The analysis of the uncertainties bands identified the differences in precipitation between the GCM-RCM combinations as being the main control, while the differences in air temperature had a low impact on the simulation outcome. This finding stands in contrast to other studies. Milly and Dunne (2011) and Sheffield et al. (2012) for example, identified empirical approaches for the calculation of evapotranspiration as the main source for overestimation of the climate's ~~sensitivity~~ influence on hydrological processes, particularly when evapotranspiration is a function of air temperature (Clark et al., 2016; Shaw and Riha, 2011; Roderick et al., 2014). In the climate scenarios used in this study, the impact of large differences in mean annual precipitation on the simulated outputs exceeded the impact of the differences in air temperature.

The effect of the model setup, with different watershed subdivisions, on the simulation of discharge or water quality variables has been investigated in various studies (e.g. Jha et al., 2004; Momm et al., 2017; Pignotti et al., 2017). Jha et al. (2004) emphasize the greater impact of changes during the HRU definition over the defined number of subbasins, as a consequent change in the distribution of land use, soil, or topography strongly affect runoff and the nutrient budget in a catchment. The analysis of the uncertainties bands with respect to the different model setups clearly confirmed the study by Jha et al. (2004), especially in the case of the Schwechat. Nevertheless, the impact of the model setup was lower than the effect of the model parametrization by a factor of up to five in the Schwechat study and up to eight in the Raab case study. Yet, the model setup strongly affects the computation time. In the present case, where aggregated discharge and NO_3^--N loads at the catchment outlets were the variables of interest a strong focus on the model parametrization is of higher priority than the spatial distribution of the model setup. Therefore, to maintain short computation times (and at the same time to maintain the distributions of land use, soil, or topography) a model setup with a low number of subbasins without any reduction of the number of HRUs is beneficial.

The impact of parameter non-uniqueness on the simulation of hydrological and water quality variables has been demonstrated previously (e.g.; Wilby, 2005; Mehdi et al., 2018). The importance of the model parametrization for the simulation of discharge and NO_3^--N loads was confirmed in the present study as well. Large sensitivities of ~~the different model parametrizations were identified for~~ all signature measures of discharge and NO_3^--N loads to the different model parametrizations were identified . Although all selected parameter sets represented historical observations of discharge and NO_3^--N loads with a certain goodness of fit (based on defined objective criteria), the colored grouping of the uncertainty bands illustrated that the selected model parameter sets control the simulation of future discharge and NO_3^--N loads in different ways. Thus, the large ~~sensitivities~~ impact of the model parametrization and the distinctive patterns identified in the uncertainty bands suggest a great potential to further refine the model parametrization and consequently reduce simulation uncertainties with a more intensive model calibration. Additional information on the time series of observations can help to constrain the model parameters and adequately describe the relevant processes (e.g. Hrachowitz et al., 2014; Pfannerstill et al., 2017).

4.2 How to attribute subjectivity inherent in the scenarios

Scenarios always reflect subjective assumptions made by the modeler. Assumptions that are made in the scenario development however, can strongly influence a simulation and thus affects a comparison of different model inputs and their impacts on the simulation. All steps in a scenario development involve subjective assumptions and can lack plausibility (Mahmoud et al., 2009; van Vuuren et al., 2012), regardless of whether the process involves expert knowledge, the input of stakeholders in an participatory process, or an exploratory approach that extrapolates trends, these practices potentially introduce uncertainties in the definition of scenarios. Technical aspects such as how the scenario is represented in the model are also strongly biased by the modelers decision and represent an additional source of uncertainty (Mahmoud et al., 2009). The communication of the potential uncertainties inherent in the developed scenarios and the boundaries of the explanatory power of an scenario ensemble is essential for the integrity of any impact study (Mahmoud et al., 2009; Jones et al., 2014).

In the present study, several assumptions were made in the development of scenarios that are highly subjective, such as the extrapolated gradient of future land use changes, the drastic changes in future industrial emissions, and also the selection of objective criteria that define a behavioral SWAT model setup. ~~For the SA of the simulated variables the diversity of the developed scenarios is essential. Thus, scenarios~~ Scenarios must cover a broad range of possible futures and have to be adequately represented in the model setup. An explicit delineation of the implemented scenarios and their limitations is essential to clearly illustrate the limitations of an impact study's conclusions. An immanent risk in any impact study is that the model representation of a future change, or the uncertainties in a model input fail to reproduce the response of a simulated variable that would have taken place in the real environmental system. Hence, a detailed analysis of the simulation uncertainties perfectly complements a SA to identify possible shortcomings in the study setup. Attributing the uncertainty bands resulting from the simulation of an environmental variable to individual model inputs prove to be a useful visual analysis tool that gives the power to illustrate the uncertainties in a transparent way. Furthermore, the colored differentiation provides a visual guidance to judge the impacts of different implemented scenarios.

4.3 Sensitivity analysis or hydrologic storylines

The presented approach implements large samples combining scenarios for different model inputs and different model setups and parametrizations in a GSA to identify the dominant contributors of uncertainties in the simulated outputs. The utilization of SA with large sample sizes however, raises the following issues: i) compared to a standard approach to perform an impact assessment, where a few different future scenarios are implemented into a model, the computational demand of a GSA requiring hundreds or thousands of model executions is larger by several orders of magnitude. Thus, a practical implementation of the presented procedure in impact studies is questionable and a strong cooperation between research and the practitioners is essential. ii) scenarios of different model inputs are often interrelated (Mahmoud et al., 2009). A change in one model input therefore for example expects the change of another model input into one direction and makes a change into another direction unlikely. ~~The application of sampling strategies for SA usually do not account for the circumstances that one model input constrains any other model input~~ While the implementation of input dependencies, althouh challenging is feasible for continuous model inputs, for instance by a transformation of the input space (e.g., Tarantola and Mara, 2017; Mara and Tarantola, 2012), or the determination of input distribution functions (Hart and Gremaud, 2018), the dependencies of composite model inputs are usually difficult to express mathematically. To identify the dependencies between composite model inputs, expert knowledge is required to properly constrain the model input combinations and therefore complicates the implementation in approaches, such as the presented one.

Clark et al. (2016) therefore suggest to identify consistent hydrologic ~~storylines~~ story lines that result in least severe, most likely, and most severe responses of the modeled system. Such an approach would tremendously reduce the number of necessary model evaluations, but also establish consistency between the considered influencing variables. Nevertheless, the feasible combinations of influencing variables that lead to extreme or likely responses of the modeled system are hardly known a priori. Consequently, a sensitivity analysis with a constrained sampling space, to avoid infeasible combinations of influencing variables might be a pragmatic compromise.

5 Conclusions

In this study we utilized methods for GSA in environmental impact studies to identify the dominant sources of uncertainties for the simulation of environmental variables under future changing conditions. In two Austrian case studies for the rivers Schwechat and Raab, we simulated the river discharge and the NO_3^--N loads from the catchments under the condition of future changes in climate, land use, and emissions from urban and industrial point sources implementing different SWAT model setups with various model parametrizations.

Both case studies identified climate change and the model parametrization to be the most important (~~sensitive~~influential) model inputs for the simulation of discharge and NO_3^--N loads, based on performing a GSA and on the resulting analysis of signature measures of discharge and NO_3^--N loads (quantiles of discharge and NO_3^--N loads, seasonal mean discharge and seasonal sums of NO_3^--N loads and NO_3^--N concentrations for discharge quantiles). The impact of the model setup on simulated variables of discharge and NO3-N loads was found to be considerably lower than the impact of the model parametrization ~~by a factor of up to 5~~ for the Schwechat and ~~by a factor of up to 8~~ even more distinct for the Raab. The impact of the implemented scenarios for land use and municipal point source emissions were negligible for all analyzed signature measures. Because of a large leather industry in the Raab catchment, the future development of industrial emission in the Raab catchment was found to be relevant for low NO_3^--N loads and NO_3^--N concentrations during low discharge.

Accompanying the GSA, a detailed analysis of the simulation uncertainties provided additional insights on how the uncertainties in the model inputs control simulated discharge and NO_3^--N loads. The visualizations we developed proved to be an effective tool to identify the relevant properties of the model inputs that control the simulation uncertainties and provide insight how individual realizations of a model input can affect the simulations. In the climate simulations, we found the precipitation to dominate the simulation outputs, rather than changes in air temperature. Although the impact of the model setup on the simulation of discharge and NO_3^--N loads was low, the visual analysis of the uncertainty bands illustrated that the HRU definition is an important step in the model setup. The use of the full set of HRUs was identified as the preferred setup in the two case studies. In contrast the effect of using different numbers of subbasins in the model setup was low for the simulation of discharge and NO_3^--N loads at the catchment outlets.

The drawn conclusions are the result of specific conditions and the assumptions made for each individual catchment in the two case studies. The conclusions cannot be extrapolated with ease to other catchments. Nevertheless, the presented work provides an approach to identify and analyze the dominant sources of simulation uncertainties in environmental impact studies that can easily be generalized and that can act as a template for further impact studies. The analyses advocate for a stronger focus on the communication of uncertainties in model simulation and their sources ~~. The~~ in environmental impact studies. Although a variety of tools to perform SA are available for different programming languages (e.g., Pianosi and Wagener, 2015; Reusser, 2015; Iooss et , the main constraint for a practical application ~~however, remains the lack of tools that allow the practitioners access to~~ remains the development of a comprehensive set of discrete input realizations, the computational costs of such analysis, and the lack if straight forward methods to implement composite inputs into SA. This might detain the practical application of such methods. ~~As a consequence~~To facilitate the implementation of composite model inputs in SA, we plan to implement the demonstrated procedures and tools for visualization into a user friendly programming environment.

Appendix A: Supplementary figures and tables

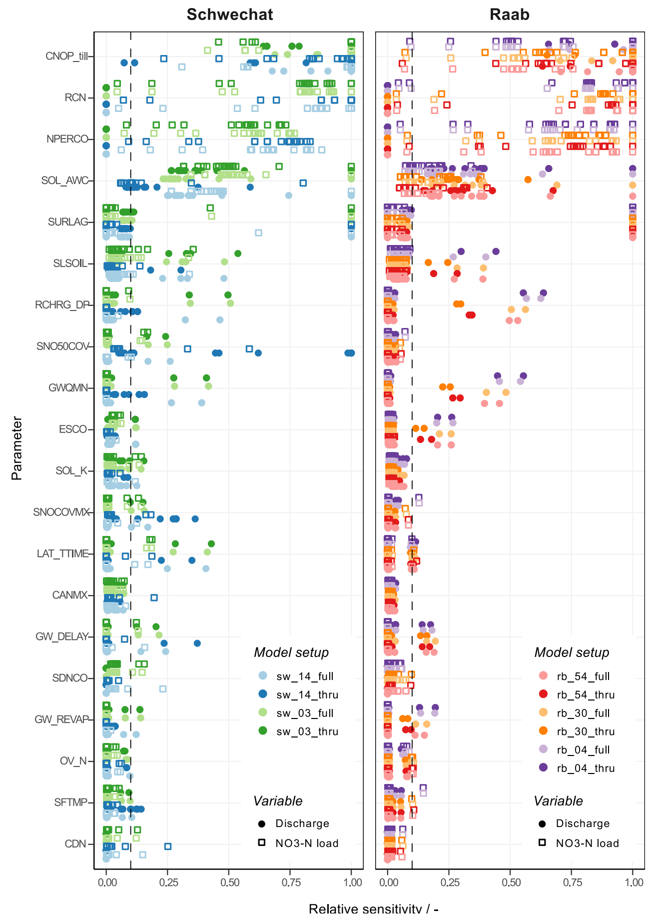

Figure A1. Identification of the ~~sensitive~~ influential SWAT model parameters for the case studies Schwechat (left) and Raab (right). The y-axis illustrates model parameters that showed ~~a sensitivity to~~ an impact on at least one of the analyzed objective ~~criterion~~criteria. The x-axis shows the relative ~~sensitivity~~ sensitivities of ~~the model parameters~~ analyzed objective criteria (in relation to the most ~~sensitive~~ influential parameter for an objective criterion). The colors indicate the different SWAT model setups. The circles show the sensitivities for objective criteria related to discharge, while the hollow squares show parameter sensitivities for $NO_3^- - N$ loads. The dashed line indicates the 0.1 value of relative sensitivity. A parameter is considered to be sensitive if it ~~showed~~ resulted in a relative sensitivity above this threshold for the objective criteria.

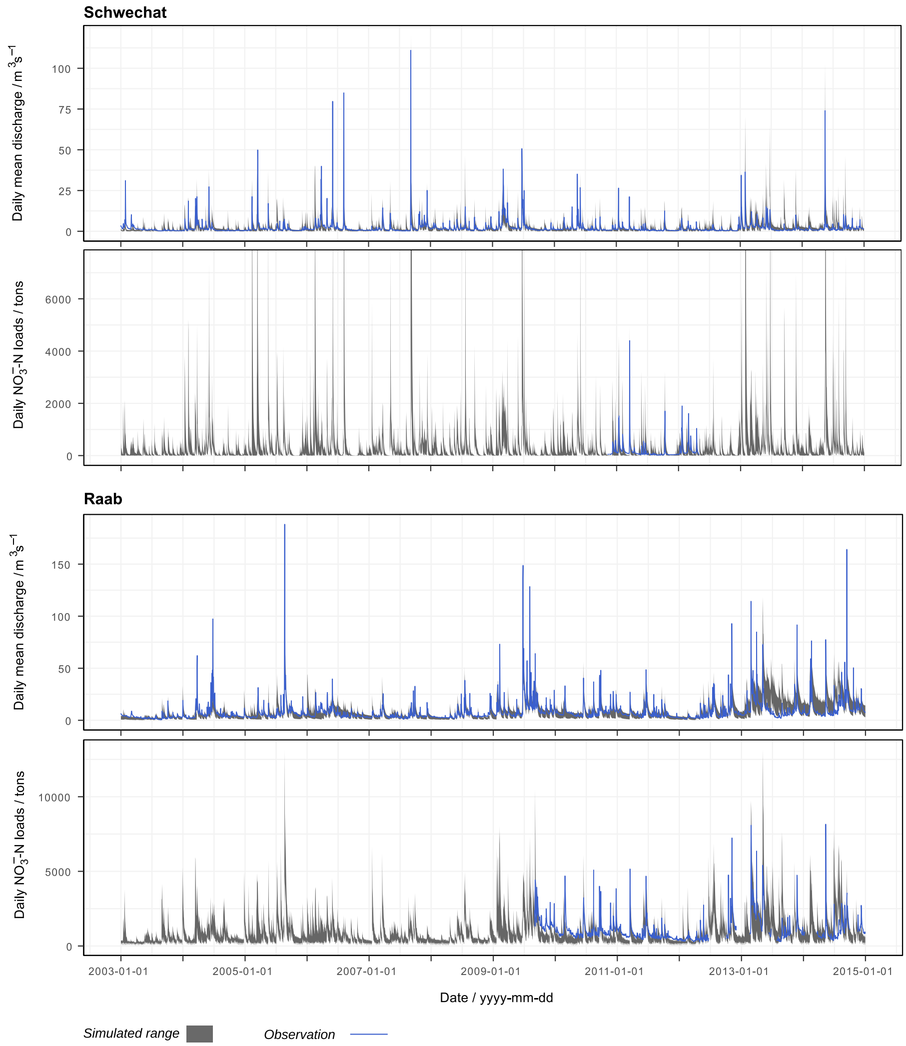

Figure A2. Simulated time series of daily mean discharge and daily $NO_3^- - N$ loads for the Schwechat (top) and the Raab (bottom) catchments for the time period 2003 to 2015. The gray bands show the ranges simulated using the selected model parameter sets with the different SWAT model setups. The blue solid lines indicate available observations of discharge and $NO_3^- - N$ loads for the respective time periods.

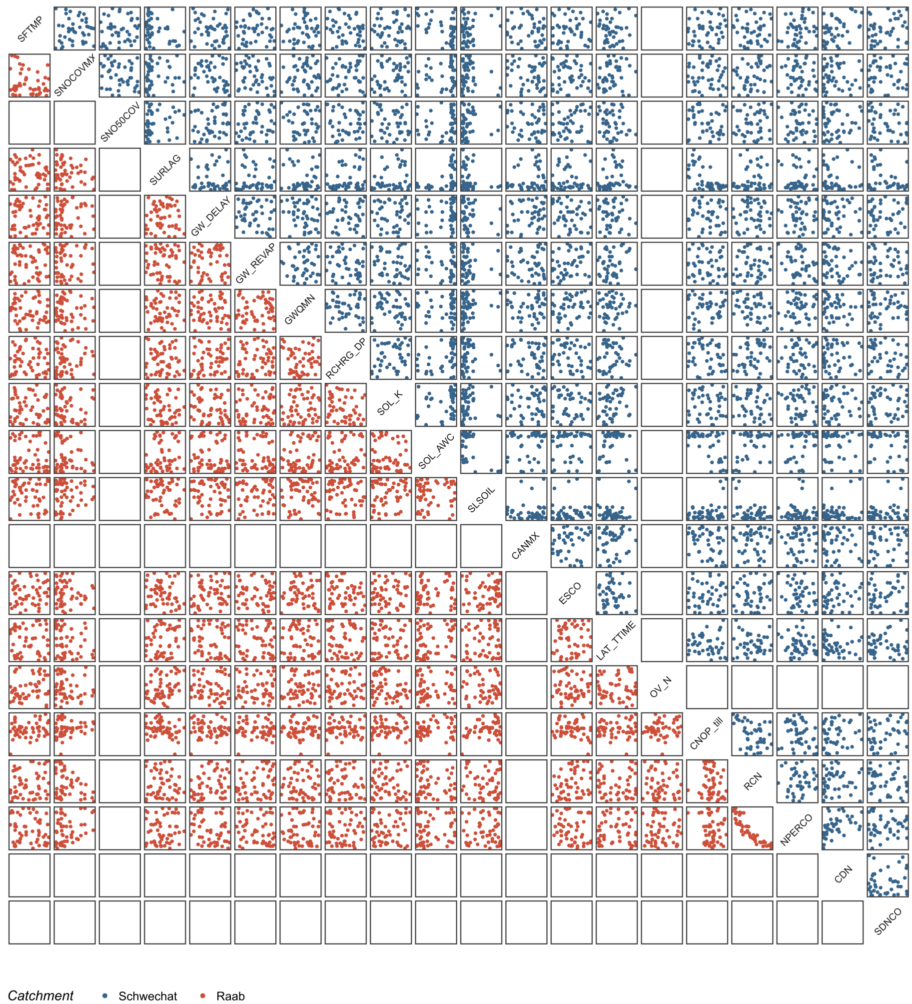

Figure A3. Parallel coordinate plot of the 43 and 52 behavioral SWAT model parameter combinations that were used with the model setups of the Schwechat and the Raab, respectively. Each panel illustrates the interaction of two model parameters. The parameter combinations for the Schwechat are illustrated in red (below the diagonal) and the combinations for the Raab are given in blue (above the diagonal). The x and y axes of each panel show the range of the respective parameter plotted along the x or y dimension. The corresponding parameter ranges for all illustrated parameters are provided in Table A2.

Table A1. Sensitive SWAT model parameters for the model setups of the Schwechat and the Raab.

		Influential for discharge		Influential for NO_3^--N loads	
Parameter	Description	Schwechat	Raab	Schwechat	Raab
SFTMP	Snowfall temperature (°C)	X			X
SNOCOVMX	Minimum snow water content that corresponds to 100% snow cover	X		X	X
SNO50COV	Snow water equivalent that corresponds to 50% snow cover	X		X	
SURLAG	Surface runoff lag time (h)	X	X	X	X
GW_DELAY	Groundwater delay (d)	X	X	X	
GW_REVAP	Groundwater revaporation coefficient	X	X		
GWQMN	Threshold depth of water in shallow aquifer for return flow (mm)	X	X		
RCHRG_DP	Deep aquifer percolation fraction	X	X		
SOL_K	Saturated hydraulic conductivity (mm/h)	X		X	
SOL_AWC	Available water capacity of the soil layer	X	X	X	
SLSOIL	Slope length for lateral subsurface flow	X	X	X	X
CANMX	Maximum canopy storage			X	
ESCO	Soil evaporation compensation factor	X	X		
LAT_TTIME	Lateral flow travel time	X	X	X	X
OV_N	Manning's n-value for overland flow		X		X
CNOP_till	SCS runoff curve number for the tillage operation	X	X	X	X
RCN	Concentration of nitrogen in rainfall			X	X
NPERCO	Nitrogen percolation coefficient			X	X
CDN	Denitrification exponential rate coefficient			X	
SDNCO	Denitrification threshold water content			X	

Table A2. Ranges of parameter changes for the behavioral model parameter sets. The type of change indicates whether a model parameter was replaced by absolute values, altered by adding an absolute to the initial parameter value, or changed by a relative fraction of the initial parameter value. The initial ranges of parameter changes and the ranges of parameter ranges of the behavioral parameter combinations in the model setups of the Schwechat and the Raab are shown.

		Range of parameter change		
Parameter	Type of change	Initial range	Schwechat	Raab
SFTMP	replace value	[-1.00, 1.00]	[-0.69, 0.93]	[-0.98, 0.88]
SNOCOVMX	replace value	[100.0, 500.0]	[0.9, 177.0]	[100.8, 447.5]
SNO50COV	replace value	[0.20, 0.50]	[0.21, 0.49]	
SURLAG	replace value	[0.00, 18.00]	[0.02, 0.99]	[0.01, 0.10]
GW_DELAY	replace value	[0.0, 300.0]	[5.5, 25.0]	[2.1, 283.3]
GW_REVAP	replace value	[0.02, 0.20]	[0.05, 0.15]	[0.02, 0.20]
GWQMN	replace value	[0, 3000]	[567, 2472]	[109, 2925]
RCHRG_DP	replace value	[0.01, 1.00]	[0.31, 0.69]	[0.13, 0.97]
SOL_K	relative change	[-0.90, 10.00]	[0.00, 0.97]	[-0.79, 9.76]
SOL_AWC	relative change	[-0.90, 2.00]	[-0.86, 1.49]	[0.01, 1.98]
SLSOIL	replace value	[0.0, 150.0]	[0.9, 27.6]	[14.7, 148.2]
CANMX	relative change	[-0.90, 2.50]	[0.34, 2.40]	
ESCO	replace value	[0.00, 0.90]	[0.05, 0.9]	[0.05, 0.89]
LAT_TTIME	replace value	[0.0, 180.0]	[0.8, 6.8]	[5.5, 176.3]
OV_N	absolute change	[-0.09, 0.60]		[0.07, 0.58]
CNOP_till	relative change	[-0.20, 0.10]	[-0.19, -0.06]	[-0.18, 0.01]
RCN	replace value	[2.00, 10.00]	[5.05, 9.97]	[2.30, 8.45]
NPERCO	replace value	[0.00, 1.00]	[0.24, 0.99]	[0.18, 0.7]
CDN	replace value	[0.00, 1.50]	[0.01, 1.44]	
SDNCO	replace value	[0.00, 0.50]	[0.02, 0.49]	

Table A3. Area and percentage of the land uses in the Schwechat catchment. The land use groups are the respective land uses shown in Fig. 1 and are derived from CORINE. With a higher thematic resolution the land uses that were implemented in the SWAT models are listed providing their areas and their percentages in the catchment.

Land use group	CORINE Level 3	Land use	SWAT Land use	Area / ha	Percentage / %
Urban/Industrial	11X, 14X	Urban medium ~~densitiy~~ density	URMD	154.2	0.6
	11X, 14X	Urban medium/low density	URML	2388.3	8.7
	12X	Industrial	UIDU	209.5	0.8
Agriculture, Complex Cultiv.	221, 222, 242	Winter wheat, winter grains	WWHT	667.6	2.4
		Spring wheat, summer grains	SWHT	317.8	1.2
		Corn, Maize	CORN	111.5	0.4
		Vegetables grouped	SGBT	74.1	0.3
		Sunflower	SUNF	30.0	0.1
		Soybean	SOYB	19.7	0.1
		Orchard, Fruit trees	ORCD	25.6	0.1
		Vineyard	GRAP	699.5	2.5
Grassland, Complex Cultiv.	231, 242	Pasture, extensive use	FESC	2406.6	8.8
		Pasture, intensive use	FESI	762.9	2.8
		Alfalfa, clover, etc.	ALFA	400.7	1.5
Deciduous forest	311	Forest, deciduous	FRSD	12941.3	47.1
Coniferous forest	312	Forest evergreen	FRSE	1152.2	4.2
Mixed forest	312	Forest, mixed	FRST	5138.4	18.7
				27499.9	100.0

Table A4. Area and percentage of the land uses in the Raab catchment. The land use groups are the respective land uses shown in Fig. 1 and are derived from CORINE. With a higher thematic resolution the land uses that were implemented in the SWAT models are listed providing their areas and their percentages in the catchment.

Land use group	CORINE Level 3	Land use	SWAT Land use	Area / ha	Percentage / %
Urban/Industrial	11X, 14X	Urban medium/low density	URML	11850.8	12.0
Agriculture, Complex Cultivation	221, 222, 242	Corn, Maize	CORN	11982.5	12.1
		Oil seed pumpkin	OELK	3171.1	3.2
		Vegetables grouped	SGBT	3035.9	3.1
		Winter wheat, winter grains	WWHT	1855.6	1.9
		Spring wheat, summer grains	WWHT	981.9	1.0
		Soybean	SOYB	445.9	0.5
		Orchard, fruit trees	ORCD	3036.1	3.1
Grassland, Complex Cultivation	231, 242	Pasture, extensive use	FESC	11635.7	11.8
		Pasture, intensive use	FESI	8474.0	8.6
		Alfalfa, clover, etc.	ALFA	598.0	0.6
Deciduous forest	311	Forest, deciduous	FRSD	15379.4	15.6
Coniferous forest	312	Forest evergreen	FRSE	7773.2	7.9
Mixed forest	312	Forest, mixed	FRST	18540.2	18.8
Waterbodies	41X	Wetlands, mixed	WETL	55.4	0.1
				98815.9	100.0

Table A5. Transformations of land uses (LUSE) in the implemented land use scenarios at the Schwechat and the Raab.

	"business-as-usual"			"extensive"	
From LUSE	To LUSE	Change %/ha	From LUSE	To LUSE	Change %/ha
Schwechat:					
Urban, light	Urban, dense	10 / 239	Winter wheat	Ext. pasture	27.5 / 184
Ext. pasture	Urban, light	15 / 361	Winter wheat	Legumes	27.5 / 184
Ext. pasture	Winter wheat	20 / 481			
Raab:					
Ext. pasture	Corn	75 / 8726	Corn	Ext. pasture	27.5 / 3595
Sugar beet	Corn	80 / 2429	Corn	Legumes	27.5 / 3595
Legumes	Corn	70 / 419			
Winter wheat	Corn	30 / 557			

Table A6. Municipal point source emissions and changes in the emissions due to different population growth scenarios in the Schwechat and the Raab catchments.

District	Scenario BAU/BPS			Scenario OROK		
	Change / %	Population	NO_3^--N / $kg \cdot yr^{-1}$	Change / %	Population	NO_3^--N / $kg \cdot yr^{-1}$
Baden (Schwechat)	0.0	32058	39842	+32.0	42317	52591
Total Schwechat	0.0	32058	39842	+32.0	42317	52591
Weiz (Raab)	+7.7	56982	44918	-2.0	51529	40872
Südoststeiermark (Raab)	+2.3	32296	16537	-20.4	25117	12868
Total Raab	+5.7	89278	61455	-8.7	76646	53740

Table A7. Industrial point source emissions and implemented changes in the emissions at the Raab due to increase in production or relocation of the dominant leather producer.

Industrial emitter	Relocation of leather industry		Increase in production	
	Change / %	NO_3^--N / $kg \cdot yr^{-1}$	Change / %	NO_3^--N / $kg \cdot yr^{-1}$
Agrana Fruit Austria GmbH	0.0	1029	0.0	1029
BOXMARK Leder/Feldbach	-100.0	0	30.0	88257
BOXMARK Leder/Jennersdorf	-100.0	0	30.0	36442
Fleischhof Raabtal GmbH	0.0	292	0.0	292
Johann Titz GmbH	0.0	3774	0.0	3774
WOLLSDORF Leder	0.0	26572	0.0	26572
Total	-75.20	31667	22.6	156366

Table A8. GCM-RCM combinations implemented in the study with their long-term mean annual precipitation sums and long-term mean annual temperatures for the Schwechat and the Raab.

	Schwechat		Raab	
Model	P / $mm \cdot yr^{-1}$	T / °C	P / $mm \cdot yr^{-1}$	T / °C
EUR-11_CNRM-CERFACS-CNRM-CM5_RCP45_CLMcom-CCLM4-8-17	845.6	10.5	1103.0	12.4
EUR-11_CNRM-CERFACS-CNRM-CM5_RCP85_CLMcom-CCLM4-8-17	828.7	11.6	1075.6	13.7
EUR-11_CNRM-CERFACS-CNRM-CM5_RCP45_SMHI-RCA4	911.9	10.9	1118.0	12.6
EUR-11_CNRM-CERFACS-CNRM-CM5_RCP85_SMHI-RCA4	943.8	12.4	1091.0	14.4
EUR-11_ICHEC-EC-EARTH_RCP45_CLMcom-CCLM4-8-17	813.3	10.6	967.0	12.5
EUR-11_ICHEC-EC-EARTH_RCP85_CLMcom-CCLM4-8-17	809.2	12.1	941.5	14.4
EUR-11_ICHEC-EC-EARTH_RCP45_SMHI-RCA4	915.8	11.2	1018.4	12.9
EUR-11_ICHEC-EC-EARTH_RCP85_SMHI-RCA4	939.7	12.9	1036.1	15.1
EUR-11_ICHEC-EC-EARTH_RCP45_KNMI-RACMO22E	772.7	10.9	965.0	12.6
EUR-11_ICHEC-EC-EARTH_RCP85_KNMI-RACMO22E	779.0	12.6	925.6	14.6
EUR-11_ICHEC-EC-EARTH_RCP45_DMI-HIRHAM5	925.8	10.4	962.8	12.4
EUR-11_ICHEC-EC-EARTH_RCP85_DMI-HIRHAM5	912.9	12.1	976.8	14.4
EUR-11_IPSL-IPSL-CM5A-MR_RCP45_IPSL-INERIS-WRF331F	907.2	10.2	1046.7	13.0
EUR-11_IPSL-IPSL-CM5A-MR_RCP85_IPSL-INERIS-WRF331F	996.2	11.6	1202.2	14.6
EUR-11_IPSL-IPSL-CM5A-MR_RCP45_SMHI-RCA4	899.8	11.7	1076.8	13.7
EUR-11_IPSL-IPSL-CM5A-MR_RCP85_SMHI-RCA4	934.6	13.5	1217.3	15.9
EUR-11_MPI-M-MPI-ESM-LR_RCP45_CLMcom-CCLM4-8-17	839.1	11.5	960.5	13.6
EUR-11_MPI-M-MPI-ESM-LR_RCP85_CLMcom-CCLM4-8-17	867.9	13.3	913.2	15.7
EUR-11_MOHC-HadGEM2-ES_RCP45_SMHI-RCA4	974.4	11.6	1108.5	13.6
EUR-11_MOHC-HadGEM2-ES_RCP85_SMHI-RCA4	945.0	13.6	1117.4	15.9
EUR-11_MOHC-HadGEM2-ES_RCP45_SMHI-RCA4	781.1	10.2	940.3	12.2
EUR-11_MOHC-HadGEM2-ES_RCP85_SMHI-RCA4	813.2	12.0	1021.4	14.3

Author contributions. Christoph Schürz, Karsten Schulz, and Bano Mehdi developed the study framework and prepared the manuscript. Christoph Schürz designed and performed all analyses illustrated in the paper. Bano Mehdi and Christoph Schürz acquired all SWAT model input data, set up the models, and developed the land use change scenarios, Brigitta Hollosi and Christoph Matulla developed the future climate change scenarios, and Alexander Pressl and Thomas Ertl calculated present wastewater emissions and developed the future municipal and industrial emission scenarios.

Competing interests. The authors declare no competing interests.

Acknowledgements. This work is a result from the project UnLoadC^3 (Project No.: KR13AC6K11021) funded by the Austrian Climate and Energy Fund in the 6th call of the ACRP program line. We ~~greatfully~~ gratefully obtained time series data of NO_3^--N concentration from Stefan Schuster (TBS Water Consult) and Roland Fuiko (IWR TU Wien) who manage the Raab monitoring data at the Stations Takern II and Neumarkt a.d. Raab. We want to thank Francesca Pianosi, Björn Guse, and the two anonymous reviewers for their detailed comments that helped to substantially improve the paper.

References

Amt d. Stmk LReg: Regionale Bevölkerungsprognose Steiermark 2015/16 -Bundesland, Bezirke und Gemeinden, Tech. rep., Graz, Austria, http://docplayer.org/32447223-Regionale-bevoelkerungsprognose-steiermark-2015-16-bundesland-bezirke-und-gemeinden.html, 2016.

Anderson, B., Borgonovo, E., Galeotti, M., and Roson, R.: Uncertainty in climate change modeling: can global sensitivity analysis be of help?, Risk analysis, 34, 271–293, https://doi.org/10.1111/risa.12117, 2014.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large area hydrologic modeling and assessment part I: model development, Journal of the American Water Resources Association, 34, 73–89, https://doi.org/10.1111/j.1752-1688.1998.tb05961.x, 1998.

Baroni, G. and Tarantola, S.: A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study, Environmental Modelling & Software, 51, 26–34, https://doi.org/10.1016/j.envsoft.2013.09.022, 2014.

Beven, K.: The limits of splitting: Hydrology, Science of the Total Environment, 183, 89–97, https://doi.org/10.1016/0048-9697(95)04964-9, 1996.

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

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology, 249, 11 – 29, https://doi.org/https://doi.org/10.1016/S0022-1694(01)00421-8, 2001.

BGBl. 1996/210: Verordnung des Bundesministers für Land- und Forstwirtschaft über die Begrenzung von Abwasseremissionen aus Abwasserreinigungsanlagen für Siedlungsgebiete (1. AEV für kommunales Abwasser), 1996.

BGBl. II 2006/96: Qualitätszielverordnung Chemie Oberflächen-gewässer (QZV Chemie OG), 2006.

BGBl. II 2010/99: Qualitätszielverordnung Ökologie Oberflächen-gewässer (QZV Ökologie OG), 2010.

BGBl. II Nr. 10/1999: Verordnung des Bundesministers für Land- und Forstwirtschaft über die Begrenzung von Abwasseremissionen aus Gerbereien, Lederfabriken und Pelzzurichtereien (AEV Gerberei), 1999.

BGBl. II Nr. 12/1999: Verordnung des Bundesministers für Land- und Forstwirtschaft über die Begrenzung von Abwasseremissionen aus der Schlachtung und Fleischverarbeitung (AEV Fleischwirtschaft), 1999.

Bieger, K., Hörmann, G., and Fohrer, N.: The impact of land use change in the Xiangxi Catchment (China) on water balance and sediment transport, Regional Environmental Change, 15, 485–498, https://doi.org/10.1007/s10113-013-0429-3, 2013.

BMLFUW: IMW3: Integrierte Betrachtung eines Gewässerabschnitts auf Basis kontinuierlicher und validierter Langzeitmessreihen [Integrated Monitoring of a river section on the basis of continuous and validated long measurement time series] (In German), Tech. rep., Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Sektion VII Wasser, Vienna, 2013.

BMLFUW: Online monitoring at the Station Neumarkt/Raab at the River Raab. Operated by the TU Wien, Institut für Gewässergüte, Resourcenmanagement und Abfallwirtschaft, 2015a.

BMLFUW: Online monitoring at the Station St.Margarethen/Takern II at the River Raab. Operated by TBS Water Consult., 2015b.

Borgonovo, E., Lu, X., Plischke, E., Rakovec, O., and Hill, M. C.: Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box, Water Resources Research, pp. 7933–7950, https://doi.org/10.1002/2017WR020767, 2017.

Butler, M. P., Reed, P. M., Fisher-Vanden, K., Keller, K., and Wagener, T.: Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis, Environmental modelling & software, 59, 10–29, https://doi.org/10.1016/j.envsoft.2014.05.001, 2014.

Canty, A. and Ripley, B. D.: boot: Bootstrap R (S-Plus) Functions, r package version 1.3-20, 2017.

Chiew, F. H. and Vaze, J.: Hydrologic nonstationarity and extrapolating models to predict the future: Overview of session and proceeding, in: IAHS-AISH Proceedings and Reports, vol. 371, pp. 17–21, https://doi.org/10.5194/piahs-371-17-2015, 2015.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resources Research, 44, https://doi.org/10.1029/2007WR006735, 2008.

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., Fowler, H. J., Prudhomme, C., Arnold, J. R., and Brekke, L. D.: Characterizing Uncertainty of the Hydrologic Impacts of Climate Change, Current Climate Change Reports, 2, 55–64, https://doi.org/10.1007/s40641-016-0034-x, 2016.

Cuntz, M., Mai, J., Zink, M., Thober, S., Kumar, R., Schäfer, D., Schrön, M., Craven, J., Rakovec, O., Spieler, D., Prykhodko, V., Dalmasso, G., Musuuza, J., Langenberg, B., Attinger, S., and Samaniego, L.: Computationally inexpensive identification of noninformative model parameters by sequential screening, Water Resources Research, 51, 6417–6441, https://doi.org/10.1002/2015WR016907, 2015.

Dai, H. and Ye, M.: Variance-based global sensitivity analysis for multiple scenarios and models with implementation using sparse grid collocation, Journal of Hydrology, 528, 286–300, https://doi.org/10.1016/j.jhydrol.2015.06.034, 2015.

Dai, H., Ye, M., Walker, A. P., and Chen, X.: A new process sensitivity index to identify important system processes under process model and parametric uncertainty, Water Resources Research, 53, 3476–3490, https://doi.org/10.1002/2016WR019715, 2017.

Dile, Y. T., Daggupati, P., George, C., Srinivasan, R., and Arnold, J.: Introducing a new open source GIS user interface for the SWAT model, Environmental modelling & software, 85, 129–138, https://doi.org/10.1016/j.envsoft.2016.08.004, 2016.

Duran-Encalada, J. A., Paucar-Caceres, A., Bandala, E. R., and Wright, G. H.: The impact of global climate change on water quantity and quality: A system dynamics approach to the US–Mexican transborder region, European Journal of Operational Research, 256, 567–581, https://doi.org/10.1016/j.ejor.2016.06.016, 2017.

EEA: CORINE Land Cover 2006 raster data. Version 17 (12/2013), http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3, 2015.

Efron, B.: Better bootstrap confidence intervals, Journal of the American statistical Association, 82, 171–185, https://doi.org/10.2307/2289144, 1987.

ESRI: ArcGIS Desktop: Release 10.1, 2012.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H.: A framework to assess the realism of model structures using hydrological signatures, Hydrology and Earth System Sciences, 17, 1893–1912, https://doi.org/10.5194/hess-17-1893-2013, 2013.

Geoland.at: Digitales Geländemodell (DGM) Österreich, https://www.data.gv.at/katalog/dataset/b5de6975-417b-4320-afdb-eb2a9e2a1dbf, 2015.

Godet, M. and Roubelat, F.: Creating the future : The use and misuse of scenarios, Long Range Planning, 29, 164–171, https://doi.org/10.1016/0024-6301(96)00004-0, 1996.

Gupta, H. and Razavi, S.: Challenges and Future Outlook of Sensitivity Analysis, in: Sensitivity Analysis in Earth Observation Modelling, edited by Petropoulos, G. P. and Srivastava, P. K., chap. 20, pp. 397–415, Elsevier, 1 edn., https://doi.org/10.1016/B978-0-12-803011-0.00020-3, 2017.

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

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Guse, B., Pfannerstill, M., and Fohrer, N.: Dynamic Modelling of Land Use Change Impacts on Nitrate Loads in Rivers, Environmental Processes, 2, 575–592, https://doi.org/10.1007/s40710-015-0099-x, 2015.

Guse, B., Pfannerstill, M., Gafurov, A., Fohrer, N., and Gupta, H.: Demasking the integrated information of discharge: Advancing sensitivity analysis to consider different hydrological components and their rates of change, Water Resources Research, 52, 8724–8743, https://doi.org/10.1002/2016WR018894, 2016a.

Guse, B., Pfannerstill, M., Strauch, M., Reusser, D. E., Lüdtke, S., Volk, M., Gupta, H., and Fohrer, N.: On characterizing the temporal dominance patterns of model parameters and processes, Hydrological Processes, 30, 2255–2270, https://doi.org/10.1002/hyp.10764, 2016b.

Haas, M. B., Guse, B., Pfannerstill, M., and Fohrer, N.: A joined multi-metric calibration of river discharge and nitrate loads with different performance measures, Journal of Hydrology, 536, 534–545, https://doi.org/10.1016/j.jhydrol.2016.03.001, 2016.

Haghnegahdar, A. and Razavi, S.: Insights into sensitivity analysis of earth and environmental systems models: On the impact of parameter perturbation scale, Environmental Modelling & Software, 95, 115–131, https://doi.org/10.1016/j.envsoft.2017.03.031, 2017.

Haghnegahdar, A., Razavi, S., Yassin, F., and Wheater, H.: Multi-criteria sensitivity analysis as a diagnostic tool for understanding model behavior and characterizing model uncertainty, Hydrological Processes, 31, 4462–4476, https://doi.org/10.1002/hyp.11358, 2017.

Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B., and Gruber, C.: The Integrated Nowcasting through Comprehensive Analysis (INCA) System and Its Validation over the Eastern Alpine Region, Weather and Forecasting, 26, 166–183, https://doi.org/10.1175/2010WAF2222451.1, 2011.

Hart, J. and Gremaud, P.: Robustness of the Sobol'indices to distributional uncertainty, arXiv preprint arXiv:1803.11249v3, 2018.

Hartigan, J. A. and Wong, M. A.: Algorithm AS 136: A K-Means Clustering Algorithm, Applied Statistics, 28, 100, https://doi.org/10.2307/2346830, 1979.

Haslinger, K., Anders, I., and Hofstätter, M.: Regional climate modelling over complex terrain: an evaluation study of COSMO-CLM hindcast model runs for the Greater Alpine Region, Climate Dynamics, 40, 511–529, https://doi.org/10.1007/s00382-012-1452-7, 2013.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F.: A trend-preserving bias correction - The ISI-MIP approach, Earth System Dynamics, 4, 219–236, https://doi.org/10.5194/esd-4-219-2013, 2013.

Hengl, T., De Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLoS one, 12, 1–40, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hiebl, J. and Frei, C.: Daily temperature grids for Austria since 1961—concept, creation and applicability, Theoretical and Applied Climatology, 124, 161–178, https://doi.org/10.1007/s00704-015-1411-4, 2016.

Hinkley, D. V.: Bootstrap methods, Journal of the Royal Statistical Society. Series B (Methodological), pp. 321–337, 1988.

Hofstätter, M., Ganekind, M., and Hiebl, J.: GPARD-6: A new 60-year gridded precipitation dataset for Austria based on daily rain gauge measurements, in: DACH 2013 - Deutsch-Österreichisch-Schweizerische Meteorologen-Tagung, Innsbruck, Austria, 2013.

Houska, T., Kraft, P., Chamorro-Chavez, A., and Breuer, L.: SPOTting model parameters using a ready-made python package, PloS one, 10, e0145 180, https://doi.org/10.1371/journal.pone.0145180, 2015.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H., and Gascuel-Odoux, C.: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, Water Resources Research, 50, 7445–7469, https://doi.org/10.1002/2014WR015484, 2014.

Iooss, B., Janon, A., Pujol, G., with contributions from Khalid Boumhaout, Veiga, S. D., Delage, T., Fruth, J., Gilquin, L., Guillaume, J., Le Gratiet, L., Lemaitre, P., Nelson, B. L., Monari, F., Oomen, R., Ramos, B., Roustant, O., Song, E., Staum, J., Sueur, R., Touati, T., and Weber, F.: sensitivity: Global Sensitivity Analysis of Model Outputs, https://CRAN.R-project.org/package=sensitivity, r package version 1.15.1, 2018.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Regional Environmental Change, 14, 563–578, https://doi.org/10.1007/s10113-013-0499-2, 2014.

Jha, M., Gassman, P. W., Secchi, S., Gu, R., and Arnold, J.: Effect of Watershed Subdivision on SWAT Flow, Sediment, and Nutrient Predictions, Journal of the American Water Resources Association, 40, 811–825, https://doi.org/10.1111/j.1752-1688.2004.tb04460.x, 2004.

Jiménez, B. E., Oki, T., Arnell, N. W., Benito, G., Cogley, J. G., Döll, P., Jiang, T., and Mwakalila, S. S.: Freshwater Resources., in: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Field, C., Barros, V., Dokken, D., Mach, K., Mastrandrea, M., Bilir, T., Chatterjee, M., Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E., Levy, A., MacCracken, S., Mastrandrea, P., and White, L., pp. 229–269, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014.

Jones, R., Patwardhan, A., Cohen, S., Dessai, S., Lammel, A., Lempert, R., Mirza, M., and von Storch, H.: Foundations for Decision Making, in: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Field, C., Barros, V., Dokken, D., Mach, K., Mastrandrea, M., Bilir, T., Chatterjee, M., Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E., Levy, A., MacCracken, S., Mastrandrea, P., and White, L., pp. 195–228, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415379.007, 2014.

Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, Nature Climate Change, 3, 369–373, https://doi.org/10.1038/nclimate1716, 2013.

Land NÖ: Landwirtschaftliche Bildung in NÖ - Versuche, http://www.lako.at/de/versuche/?lang=de{&}a=179{&}a{_}urlname=versuche{&}versuche{_}a=1, 2015.

LGBl. Nr. 39/2015: Verordnung des Landeshauptmannes von Steiermark vom 20. Mai 2015, mit der ein Regionalprogramm zum Schutz der Grundwasserkörper Grazer Feld, Leibnitzer Feld und Unteres Murtal erlassen und Schongebiete bestimmt werden (Grundwasserschutzprogramm Graz bis B, 2015.

Mahmoud, M., Liu, Y., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R., Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., van Delden, H., Waldick, R., White, D., and Winter, L.: A formal framework for scenario development in support of environmental decision-making, Environmental Modelling and Software, 24, 798–808, https://doi.org/10.1016/j.envsoft.2008.11.010, 2009.

Mara, T. A. and Tarantola, S.: Variance-based sensitivity indices for models with dependent inputs, Reliability Engineering & System Safety, 107, 115–121, https://doi.org/10.1016/j.ress.2011.08.008, 2012.

Massmann, C. and Holzmann, H.: Analysing the Sub-processes of a Conceptual Rainfall-Runoff Model Using Information About the Parameter Sensitivity and Variance, Environmental Modeling and Assessment, 20, 41–53, https://doi.org/10.1007/s10666-014-9414-6, 2015.

Massmann, C., Wagener, T., and Holzmann, H.: A new approach to visualizing time-varying sensitivity indices for environmental model diagnostics across evaluation time-scales, Environmental Modelling & Software, 51, 190–194, https://doi.org/10.1016/J.ENVSOFT.2013.09.033, 2014.

Mehdi, B., Lehner, B., Gombault, C., Michaud, A., Beaudin, I., Sottile, M.-F., and Blondlot, A.: Simulated impacts of climate change and agricultural land use change on surface water quality with and without adaptation management strategies, Agriculture, Ecosystems & Environment, 213, 47–60, https://doi.org/10.1016/j.agee.2015.07.019, 2015a.

Mehdi, B., Ludwig, R., and Lehner, B.: Evaluating the impacts of climate change and crop land use change on streamflow, nitrates and phosphorus: A modeling study in Bavaria, Journal of Hydrology: Regional Studies, 4, 60–90, https://doi.org/10.1016/j.ejrh.2015.04.009, 2015b.

Mehdi, B., Schulz, K., Ludwig, R., Ferber, F., and Lehner, B.: Evaluating the Importance of Non-Unique Behavioural Parameter Sets on Surface Water Quality Variables under Climate Change Conditions in a Mesoscale Agricultural Watershed, Water Resources Management, 32, 619–639, https://doi.org/10.1007/s11269-017-1830-3, 2018.

Milly, P. C. D. and Dunne, K. A.: On the hydrologic adjustment of climate-model projections: The potential pitfall of potential evapotranspiration, Earth Interactions, 15, 1–14, https://doi.org/10.1175/2010EI363.1, 2011.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Climate change. Stationarity is dead: whither water management?, Science, 319, 573–574, https://doi.org/10.1126/science.1151915, 2008.

Momm, H. G., Bingner, R. L., Emilaire, R., Garbrecht, J., Wells, R. R., and Kuhnle, R. A.: Automated watershed subdivision for simulations using multi-objective optimization, Hydrological Sciences Journal, 62, 1564–1582, https://doi.org/10.1080/02626667.2017.1346794, 2017.

Moriasi, D., Arnold, J., Van Liew, M., Binger, R., Harmel, R., and Veith, T.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Transactions of the ASABE, 50, 885–900, https://doi.org/10.13031/2013.23153, 2007.

Morris, M. D.: Factorial sampling plans for preliminary computational experiments, Technometrics, 33, 161–174, 1991.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J.: The next generation of scenarios for climate change research and assessment, Nature, 463, 747–756, https://doi.org/10.1038/nature08823, 2010.

Muerth, M. J., Gauvin St-Denis, B., Ricard, S., Velázquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., and Turcotte, R.: On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, Hydrology and Earth System Sciences, 17, 1189–1204, https://doi.org/10.5194/hess-17-1189-2013, 2013.

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

Neitsch, S., Arnold, J., Kiniry, J., and Williams, J.: Soil and Water Assessment Tool Theoretical Documentation Version 2009, Tech. rep., Texas Water Resources Institute, Temple, Texas, 2011.

ÖWAV: ÖWAV-Regelblatt 25: Abwasserentsorgung in dünn besiedelten Gebieten, 2. vollständig überarbeitete Auflage, 2010.

Pfannerstill, M., Guse, B., and Fohrer, N.: Smart low flow signature metrics for an improved overall performance evaluation of hydrological models, Journal of Hydrology, 510, 447–458, https://doi.org/10.1016/j.jhydrol.2013.12.044, 2014.

Pfannerstill, M., Bieger, K., Guse, B., Bosch, D. D., Fohrer, N., and Arnold, J. G.: How to Constrain Multi-Objective Calibrations of the SWAT Model Using Water Balance Components, Journal of the American Water Resources Association, 53, 532–546, https://doi.org/10.1111/1752-1688.12524, 2017.

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

Pianosi, F. and Wagener, T.: Distribution-based sensitivity analysis from a generic input-output sample, Environmental Modelling & Software, 108, 197–207, https://doi.org/10.1016/j.envsoft.2018.07.019, 2018.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T.: Sensitivity analysis of environmental models: A systematic review with practical workflow, Environmental Modelling & Software, 79, 214–232, https://doi.org/10.1016/j.envsoft.2016.02.008, 2016.

Pignotti, G., Rathjens, H., Cibin, R., Chaubey, I., and Crawford, M.: Comparative analysis of HRU and grid-based SWAT models, Water, 9, 272, https://doi.org/10.3390/w9040272, 2017.

R Core Team: R: A language and environment for statistical computing., https://www.r-project.org/, 2017.

Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., and Uijlenhoet, R.: Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, Water Resour. Res, 50, 409–426, https://doi.org/10.1002/2013WR014063, 2014.

Razavi, S. and Gupta, H. V.: What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models, Water Resources Research, 51, 3070–3092, https://doi.org/10.1002/2014WR016527, 2015.

Razavi, S. and Gupta, H. V.: A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory, Water Resources Research, 52, 423–439, https://doi.org/10.1002/2015WR017558, 2016a.

Razavi, S. and Gupta, H. V.: A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application, Water Resources Research, 52, 440–455, https://doi.org/10.1002/2015WR017559, 2016b.

Reusser, D.: fast: Implementation of the Fourier Amplitude Sensitivity Test (FAST), https://CRAN.R-project.org/package=fast, r package version 0.64, 2015.

Riahi, K., Grübler, A., and Nakicenovic, N.: Scenarios of long-term socio-economic and environmental development under climate stabilization, Technological Forecasting and Social Change, 74, 887–935, https://doi.org/10.1016/j.techfore.2006.05.026, 2007.

Roderick, M. L., Sun, F., Lim, W. H., and Farquhar, G. D.: A general framework for understanding the response of the water cycle to global warming over land and ocean, Hydrology and Earth System Sciences, 18, 1575–1589, https://doi.org/10.5194/hess-18-1575-2014, 2014.

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Zeng, X., and De Gonçalves, L. G. G.: A fully multiple-criteria implementation of the Sobol method for parameter sensitivity analysis, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD016355, 2012.

Rounsevell, M. D. and Metzger, M. J.: Developing qualitative scenario storylines for environmental change assessment, https://doi.org/10.1002/wcc.63, 2010.

Ruzicka, K., Gabriel, O., Bletterie, U., Winkler, S., and Zessner, M.: Cause and effect relationship between foam formation and treated wastewater effluents in a transboundary river, Physics and Chemistry of the Earth, 34, 565–573, https://doi.org/10.1016/j.pce.2009.01.002, 2009.

Saltelli, A. and Annoni, P.: How to avoid a perfunctory sensitivity analysis, Environmental Modelling & Software, 25, 1508–1517, https://doi.org/10.1016/j.envsoft.2010.04.012, 2010.

Saltelli, A. and Tarantola, S.: On the Relative Importance of Input Factors in Mathematical Models, Journal of the American Statistical Association, 97, 702–709, https://doi.org/10.1198/016214502388618447, 2002.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M.: Sensitivity analysis in practice: A guide to assessing scientific models, vol. 91, John Wiley & Sons Ltd, Chichester, West Sussex, UK, 1 edn., 2004.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global Sensitivity Analysis. The Primer, John Wiley & Sons, Ltd, Chichester, UK, https://doi.org/10.1002/9780470725184, 2008.

Sarrazin, F., Pianosi, F., and Wagener, T.: Global Sensitivity Analysis of environmental models: Convergence and validation, Environmental Modelling & Software, 79, 135–152, https://doi.org/10.1016/j.envsoft.2016.02.005, 2016.

Savage, J. T. S., Pianosi, F., Bates, P., Freer, J., and Wagener, T.: Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model, Water Resources Research, 52, 9146–9163, https://doi.org/10.1002/2015WR018198, 2016.

Schönhart, M., Trautvetter, H., Parajka, J., Blaschke, A. P., Hepp, G., Kirchner, M., Mitter, H., Schmid, E., Strenn, B., and Zessner, M.: Modelled impacts of policies and climate change on land use and water quality in Austria, Land Use Policy, 76, 500–514, https://doi.org/10.1016/j.landusepol.2018.02.031, 2018.

Schulz, K., Beven, K., and Huwe, B.: Equifinality and the problem of robust calibration in nitrogen budget simulations, Soil Science Society of America Journal, 63, 1934–1941, https://doi.org/10.2136/sssaj1999.6361934x, 1999.

Shaw, S. B. and Riha, S. J.: Assessing temperature-based PET equations under a changing climate in temperate, deciduous forests, Hydrological Processes, 25, 1466–1478, https://doi.org/10.1002/hyp.7913, 2011.

Sheffield, J., Wood, E. F., and Roderick, M. L.: Little change in global drought over the past 60 years, Nature, 491, 435–438, https://doi.org/10.1038/nature11575, 2012.

Sheikholeslami, R., Razavi, S., Gupta, H. V., Becker, W., and Haghnegahdar, A.: Global sensitivity analysis for high-dimensional problems: How to objectively group factors and measure robustness and convergence while reducing computational cost, Environmental Modelling & Software, 111, 282–299, https://doi.org/doi.org/10.1016/j.envsoft.2018.09.002, 2019.

Smith, S. J. and Wigley, T. M.: Multi-gas forcing stabilization with minicam, Energy Journal, 27, 373–391, https://doi.org/10.5547/ISSN0195-6574-EJ-VolSI2006-NoSI3-19, 2006.

Sobol, I. M.: Sensitivity analysis for nonlinear mathematical models, Mathematical Modelling and Computational Experiments, 4, 407–414, https://doi.org/10.18287/0134-2452-2015-39-4-459-461., 1993.

Statistik Austria: ÖROK-Regionalprognosen 2014 - Bevölkerung, Ausführliche Tabellen zur kleinräumigen ÖROK-Prognose 2014, http://www.oerok.gv.at/, 2015a.

Statistik Austria: STATCube - Statistical Data base of the Statistik Austria: Agricultural census - Land use (not openly accessible), http://statcube.at/statistik.at/ext/statcube, 2015b.

Statistik Austria: Datenbank zur Bevölkerungsprognose 2016 - Hauptszenario, https://www.statistik.at/, 2016.

Statistik Austria: STATCube - Statistical Data base of the Statistik Austria: Agricultural and forestry holdings with arable land and their cultivated land area (not openly accessible), data.http://statcube.at/statistik.at/ext/statcube, 2017.

Strauch, M., Schweppe, R., and Schürz, C.: TopHRU: Threshold optimization for HRUs in SWAT, https://doi.org/10.5281/zenodo.154379, https://github.com/michstrauch/TopHRU/tree/v1.2, 2016.

Tarantola, S. and Mara, T. A.: Variance-based sensitivity indices of computer models with dependent inputs: The Fourier Amplitude Sensitivity Test, International Journal for Uncertainty Quantification, 7, https://doi.org/10.1615/Int.J.UncertaintyQuantification.2017020291, 2017.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of Hydrology, 456-457, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052, 2012.

Teutschbein, C. and Seibert, J.: Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions, Hydrology and Earth System Sciences, 17, 5061–5077, https://doi.org/10.5194/hess-17-5061-2013, 2013.

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., and Tóth, G.: New generation of hydraulic pedotransfer functions for Europe, European Journal of Soil Science, 66, 226–238, https://doi.org/10.1111/ejss.12192, 2015.

Tripathi, M. P., Raghuwanshi, N. S., and Rao, G. P.: Effect of watershed subdivision on simulation of water balance components, Hydrological Processes, 20, 1137–1156, https://doi.org/10.1002/hyp.5927, 2006.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J. F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K.: The representative concentration pathways: An overview, Climatic Change, 109, 5–31, https://doi.org/10.1007/s10584-011-0148-z, 2011.

van Vuuren, D. P., Kok, M. T., Girod, B., Lucas, P. L., and de Vries, B.: Scenarios in Global Environmental Assessments: Key characteristics and lessons for future use, Global Environmental Change, 22, 884–895, https://doi.org/10.1016/j.gloenvcha.2012.06.001, 2012.

Wagner, P. D., Bhallamudi, S. M., Narasimhan, B., Kumar, S., Fohrer, N., and Fiener, P.: Comparing the effects of dynamic versus static representations of land use change in hydrologic impact assessments, Environmental Modelling and Software, pp. 1–9, https://doi.org/10.1016/j.envsoft.2017.06.023, 2017.

Wilby, R. L.: Uncertainty in water resource model parameters used for climate change impact assessment, Hydrological Processes, 19, 3201–3219, https://doi.org/10.1002/hyp.5819, 2005.

Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C., Main, J., and Wilks, D. S.: Statistical downscaling of general circulation model output: A comparison of methods, Water Resour. Res., 34, 2995–3008, https://doi.org/10.1029/98wr02577, 1998.

Willmott, C. J., Robeson, S. M., and Matsuura, K.: A refined index of model performance, International Journal of Climatology, 32, 2088–2094, https://doi.org/10.1002/joc.2419, 2012.

Winchell, M., Srinivasan, R., Di Luzio, M., and Arnold, J. G.: ArcSWAT 2012.10.19 Interface for SWAT2012, http://swat.tamu.edu/software/arcswat/, 2015.

Wise, M., Calvin, K., Thomson, A., Clarke, L., Bond-Lamberty, B., Sands, R., Smith, S. J., Janetos, A., and Edmonds, J.: Implications of Limiting CO2 Concentrations for Land Use and Energy, Science, 324, 1183–1186, https://doi.org/10.1126/science.1168475, 2009.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, Climatic Change, 62, 189–216, https://doi.org/10.1023/B:CLIM.0000013685.99609.9e, 2004.

Yates, D. N., Miller, K. A., Wilby, R. L., and Kaatz, L.: Decision-centric adaptation appraisal for water management across Colorado's Continental Divide, Climate Risk Management, 10, 35–50, https://doi.org/10.1016/j.crm.2015.06.001, 2015.

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

Zorita, E. and Von Storch, H.: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods, Journal of Climate, 12, 2474–2489, https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2, 1999.