Author's response

Dear Prof. Dr. Nadia Ursino,

We thank you for the opportunity to improve our manuscript entitled "Exploring river–aquifer interactions and hydrological system response using baseflow separation, impulse response modelling and time series analysis in three temperate lowland catchments (HESS-2021-422)". We have addressed all the comments from the two referees in the revised manuscript.

We have reduced the length of the original manuscript from 45 to 36 pages and listed the major changes in the revised manuscript below.

Major revisions:

- Improved the description of the hydrologic data, lithological and hydrogeological settings, the natural state of the catchments in the "Study area" section
- Integrated the previous "Data collection" and "Data cleaning" sections into "Data collection" and added detailed description of the groundwater level observations
- Shortened the "Data imputation", "Model overview" of "Impulse response modelling", and removed the "Cross correlation analysis" in the "Methodology" section
- Removed previous Fig. 5, 6, 10, 12, 15, Table 2 and the relevant text in the original manuscript
- Improved visualization and annotations of all figures, and avoided the verbal explanations in the captions, e.g. Fig.2, Fig.5–7, Fig. 12–17
- Adjusted the "Results", "Discussion" and "Conclusions" sections according to the methodology, figure changes as listed above

In the following sections, we have listed the detailed changes (in blue) made to the original manuscript based on the comments (in *italic*). The page and line numbers (e.g. **P2, L26**) correspond to the revised manuscript (see the supplement).

We are looking forward to the further assessment.

On behalf of all authors,

With best regards,

Min Lu

Corresponding author

Response to Anonymous Referee #1

1. General comments

The study looks in details at the hydrological interactions between lowland rivers and shallow aquifers in three different catchments in Belgium. In their analysis, the authors use a combined approach of baseflow separation, impulse response modeling and time series analysis over a 30-year period. Overall, the paper is well-written and the results are insightful and very useful to the hydrology community. I think the paper should be published after addressing the following minor/technical comments:

- We thank the Anonymous Referee #1 for his/her useful comments and suggestions to help improve the quality of the manuscript. We have addressed all the comments in the revised manuscript, see the detailed responses to specific suggestions below.

2. Specific comments

- Looking at Fig. 11(c), I can see a slight overestimation of the groundwater levels prior to 2000. Can the impulse response parameters be tuned to improve the fit to the data? Also the seasons, as described in the text, are not clear on the figure. Consider zooming in.

- As mentioned in the previous response to Anonymous Referee #1, if we split the 30-year study period into smaller sub-periods, for example, every 10-year period, and optimize one set of impulse response parameters for each sub-period, the fitting to the observed level time series could be improved. But this approach deviates from the scope of this study, which assumes that the hydrological system has not evolved drastically within the whole 30-year time frame (1990 - 2020) and uses one set of impulse response parameters covering the whole period. The model residuals and deviations from the simulated time series may reflect effects of human interferences.

- We agree with you that Fig. 8 (originally Fig. 11) and the previously described text did not match well. We improved Fig. 8 **(P16)** by adding minor gridlines on a yearly basis for helping visualize the low points, and adjusted the text from "…exhibit very low levels at the end of summer or begin of autumn…" to "…exhibit very low levels in the middle or second half of the year…" **(P16, L349)** for clarification.

- Section 4.2.2: I am interested to look at the entire eigenvalue spectrums. I understand that the first mode at the 3 different sites dominate the rest of the modes. Does the spectrum die after the first leading modes? I expect the leading modes to change if there is some sort of an extreme rainfall event. It might be helpful to comment and discuss this further in the text.

- Yes, the spectrum dies after the first mode. To make this point clear, we improved Fig. 11 **(P20)** by adding the percentage of variances of the first five principal components in each catchment. Figure 11 shows clearly the trend of spectrum signals (Fig. 11b, d, f). Since this study mainly focuses on an overall catchment behavior for a broader time scale of 30 years, we did not assess the impacts on the mode change due to extreme rainfall events, which usually take places at much finer temporal scales such as a few hours or days. The leading modes may change due to extreme rainfall events if checking at finer temporal scales. However, this assessment is out of the research scope and focuses.

- Do you expect the BFI estimates in these groundwater-dominated lowland catchments to change if the precipitation regimes change?

- Yes, we expect the BFI will change if the precipitation regimes change. For instance, if summers are becoming drier with less precipitation (e.g. dry springs/summers observed between 2017 – 2020 in western Europe), the rivers are at risk of drying up. Expected groundwater discharge to rivers will also be limited due to lowered groundwater level.

- We think that the magnitude of impacts on the BFI due to the precipitation regime change also depends on the temporal scale of the regime change. If the precipitation regime changes over a longer period (e.g. 10 years, or even longer), the impacts on BFI will also demonstrate a long-term trend. On the contrary, if the regime change just happens for a few years, its impacts on BFI will be short-term since the river-aquifer interactions in lowland environment are very robust.

- The figures could be annotated better. For instance, in Fig. 13 does not label the black and the gray curves.

- We agree with the suggestion. We improved Fig. 9 (originally Fig. 13) by plotting the IRF curves all in black and removed the previous transparency setting **(P17)**. In this way, there is no confusion about black/grey color differences. Following your suggestions, we also improved annotations and visualization of all other figures, and avoided the verbal explanations in the captions. For instance, we adjusted the left y axis and the colors for presenting air temperature curves in Fig. 2 **(P4)**. Now it is easier to distinguish the air temperature and precipitation without specifying the plot types in the caption. We also added minor gridlines or annotations in Fig. 5–7 **(P13–15)** and Fig. 12–17 **(P22–27)** for better visualization. Other detailed adjustments on figures can be found in the revised manuscript with tracked changes.

Response to Prof. Dr. Franklin Schwartz

1. General comments

The paper by Lu et al examines surface water groundwater interactions in lowland catchments in Belgium. The purpose of the study was to fill a gap in knowledge related to these catchments (Line 61). Their approach relied on an impulse response modeling to establish baseflow from knowledge of water table fluctuations. Those baseflow estimates would then be employed to evaluate methods of hydrograph separation and to learn something about the hydrology of these lowland basins.

- We appreciate the helpful comments and suggestions for improving our manuscript. We revised the manuscript and addressed all the comments, see the detailed point-to-point responses below.

2. Specific comments

Study Design

Intuitively, I question the motivations for study, discussed in the introduction. There are many different kinds of watersheds worldwide and it is not clear why the knowledge gap in this case was worthy of *the time spent. Another question is the apparent need for another study designed to evaluate the efficacy of various baseflow separation techniques. The paper itself identified the key problem (Line 216) "Limitations for these hydrograph separation methods are their intrinsic difficulty to validate the separated baseflow and the lack of any representation of the physical processes of the river-aquifer exchange". This problem is well known and has been widely explored (e.g., Partington et al., 2012). The positive aspects mentioned in the paper "fast", "efficient", "widely used" and quantitative (line 218) really don't justify techniques know to be little more than guesses in most applications. The choices to address this problem in my opinion are to minimize this aspect of the study in the paper, demonstrate with field data that one of the approaches does work well enough to be useful, or to use a modelling approach like SWOT that might be useful.*

- As mentioned in the previous response to Prof. Dr. Franklin Schwartz, the selected sites are of interest for multiple research partners and stakeholders of the Future Floodplain Project. Studying the river-aquifer interaction can fill the missing gap and also provide input for the ecological studies in these catchments. We made this point clear in the introduction section **(P2–3, L54–59)** .

- We adjusted slightly the phrasing of the research objectives **(P3, L65–69)**. The first objective is to simulate the groundwater level response using impulse response modelling, and the second one is to simulate the groundwater inflow to rivers and compare the estimated groundwater inflow with the separated baseflow. Although we used the traditional hydrograph separation methods, we did not intend to focus on evaluating methodological differences between varying baseflow separation techniques. We removed the description such as "…we consider the digital filter Nathan and Eckhardt methods to yield more reliable results than the graphical HYSEP methods…" from the original manuscript to avoid misleading or shifting of the main research focus.

- From a broader research perspective, we have divided the river-aquifer interaction study into three main parts: (1) data-driven modelling and simulation at a catchment scale, (2) multi-method field approaches at local scales, and (3) comprehensive numerical modelling approach including both current and future scenarios (climate and land use change). This paper focuses on the first part, and including aspects of the others would make it even more lengthy.

The study in my opinion suffers from an over-reliance on theoretically based approaches. On line 48, the paper mentions several field-based approaches, but suggested that these were scale-inappropriate. There are other techniques not mentioned that have been used in other studies, e.g., isotope tracers and geochemical hydrograph separation. These of course come along with their own problems but have been applied to basins of this scale, which are small in area. The paper would be helped by field-based data/observations that could validate any of the empirical conclusions.

- We have carried out multi-method field approaches in the focus zones of the catchments (part of the whole catchments). We have collected river water and shallow groundwater samples at different seasons for geohydrochemical (major ions and cations, Radon-222) and isotopic (H2 and O18) analysis. We use heat tracer for estimating river-aquifer interactions at local scales and are currently working on it. We also explored the thermal infrared imagery technique together with Radon-222 analysis in the study sites at local scales (https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.13839). These are the main focuses of the second part of our river-aquifer interaction studies.

The decision to forego a rigorous physically based modeling, approach e.g., HydroGeoSphere, in the study design was surprising. That model was used in various baseflow application e.g., Olsthoorn et al. (2012) – an application looking at the efficacy of hydrograph separation methods (cited in the paper), and with geochemical approaches (Jones et al., 2006). Even if the study was focused on refinement of the impulse-response approach, it would have been prudent to start with a simple well constrained model-based proxy (like Olsthoorn et al. 2012).

- As part of a broader research scheme, numerical modelling of river-aquifer interactions is the main focus of the third subsection of our project. Based on our previous modelling experience, we chose MODFLOW-based numerical modelling. We are currently working on it and will compare the groundwater inflow to the river from numerical MODFLOW approach and data-driven modelling approach (of this manuscript).

What is Baseflow?

I think it is important for the authors to explain their concept of baseflow. The implicit definition in the paper is that stream baseflow is due to groundwater. For developed watersheds, baseflow is flow in the

stream between storm-runoff events. That water could be groundwater, but it also might include slow surface-water discharge from impoundments, storm-water ponds, dewatering, or discharges of treated sewage, etc (Liu et al. 2013). With this expanded definition in mind, the authors need additional field data to support their assumption that baseflow is groundwater.

Development of a watershed (farming, cities etc.) also has the potential to reduce baseflows by decreasing natural groundwater recharge due to tile drains, stormwater collection systems and fast runoff from pavements and altered land cover. To provide context for this study of low land basin these possibilities need to be explored with additional field data and observations.

- We agree it is necessary to explicitly define the concept of baseflow in our study. In the revised manuscript, we defined "baseflow, from groundwater discharge and other delayed sources" **(P8, L177–178)**. The term "baseflow response modelling" was adjusted to "groundwater inflow response modelling" **(P11, L270)** as groundwater inflow to rivers is part of the (total) baseflow. We also adjusted the corresponding terms throughout the whole manuscript where necessary to distinguish between baseflow and groundwater inflow.

- Unlike urban streams influenced by anthropogenic activities (Liu et al. 2013), the three catchments of this study are dominated by crop and meadow, and have relatively low urban coverage **(P6, L122–124)**. Human impact on these catchments was also limited during the research period. Therefore, the three catchments reflect natural conditions, which makes it feasible to implement the impulse-response modelling, especially under climatic forcing. We added more descriptions with regard to the natural state of the catchments **(P6, L124–128)**.

Questions Concerning the Data

The descriptions of these study watershed appear relatively meager in terms of hydrologic data. First, in looking at the stream hydrographs, it seemed that that discharges were unusually constrained in a *narrow range of discharges. I think that for the two smaller basins at least mean daily discharges do not provide adequate temporal resolution of discharge conditions.*

- Besides the graphical presentation of the hydrologic data (Fig. 2 and 3), we added and revised the statistical description part **(P3, L76–82)**. Regarding discharge, we chose daily discharge values over finer resolution (e.g. hourly) because we simulated the impulse response modelling over a relatively long period (30 years) and focused on the temporal evolution at a coarse temporal resolution. Also, using daily values for all three catchments helped us to compare the difference between them. For event based or finer temporal resolution studies, we agree it is better to have much finer temporal scales than daily.

In most watersheds, groundwater-level hydrographs are relatively uncommon. The record shown in the paper appears to have combined bits-and-pieces of hydrographs from different wells. But I could be mistaken. The assertion that forcing from precipitation provides a single simulated water-level fluctuation for an entire catchment is a serious simplification that has not really been appropriately justified and is not appropriate. The job of the land-surface component of hydrologic models is to redistribute water on the land surface due to topography, land cover, and soil conditions, which together provide for huge variability in local infiltration rates. Similarly, the hydrologic response of shallower wells could be substantially different than deeper wells because of local variability in hydrogeologic parameters. For example, there are no indication as to whether aquifers are fractured at shallow depth etc.

- We explored in this study the possibility to get a sense of groundwater contribution to the river from a groundwater-level hydrograph perspective, which is not so common but also can be considered as the innovative part of our study. Since the three catchments have experienced little human impact during the study period, this makes it possible to use a simple data-driven model to explore the hydrological system response under climatic forcing first, before applying a complicated numerical modelling including other influencing factors.

- As most groundwater level observations are close to the river, we used the first principal component to get a collective representation of the relative state of the shallow groundwater levels across the catchment for facilitating the application of the impulse response model for the second use case **(P11, L272–275)**. The groundwater level observations in this study are from shallow depths so they represent the water table elevation, and fractures are not of concern in these shallow unconsolidated sediments. Locally, Roer-Valley Graben related faults influence groundwater levels in the region, but in the current study areas, there is no evidence they would reach up to the shallow aquifer.

It is also noticeable that the locations of groundwater observation wells are biased to specific parts of watersheds, and to locations close to streams. Are these completed in alluvial aquifers adjacent to the stream or at what depths? Are these wells special to have water-level records, what kind of records exist etc.? How often are water levels measured in these wells.

- We tried our best to get as many groundwater level observations as possible from different monitoring agencies (e.g. Flanders and Wallonia have different monitoring networks and schemes). Since our focus is on exploring the river-aquifer interactions, the well locations close to the streams actually can provide more valuable information for the regional groundwater inflow to rivers. We have recommended that weighting scheme reflecting this might improve the approach further **(P29, L522–525)**, but consider it out of scope in the current work.

- In the original manuscript, the description of the groundwater level time series wasa bit scattered and not specific. We have improved and added more detailed description in the data collection section. We specified the sources of observations, record intervals, time series lengths, groundwater depths in the three catchments **(P7, L149–164)**.

Publication Strategy

With 23 Figures and 45 pages, this paper is overly long with several uncoordinated threads. Yet even at this size, there are major gaps in the description of the hydrogeological setting and data deficiencies that are concerning. My recommendations would be for the authors to rethink their publication strategy to create several papers with different purposes.

With modest effort, there might be a first paper to examine unique features of the hydrologic settings, especially basin morphology, elevation, land use land cover, in predicting hydrographs. It might be necessary to find an approach to reconstruct (downscale) hydrographs to improve resolution. Also, high-resolution water sampling of one storm – with specific conductance etc. together with a few groundwater samples could provide a better understanding of where inter-storm water is coming from.

A second paper might be designed to develop a more sophisticated understanding of water-level behaviors in a model system with a uniform rainfall to begin exploration of the impulse response modeling of the first link – precipitation groundwater.

Finally, a third paper might extend to modeling baseflow as you have done in this paper. But with a much better concept of how everything is working. I would also recommend that you only return to hydrograph separation with a tunable scheme that would integrate basic approaches with some kind of observational approach.

Jones et al 2006 DOI:10.1029/2006WR005416

Liu et al 2013 ISSN : 1866-6280

- We added additional descriptions on the lithostratigraphic classifications in the subsurface **(P5, L108–110, L112–115)**. Since this study is focused on the river and shallow aquifer interactions, we gave a bit more weight on describing the surficial geology.

- We agree that the paper is long. We have shortened the manuscript from 45 to 36 pages. For instance, we have integrated the previous Sect 3.1.2 to Sect 3.1.1 Data collection, removed previous Fig. 5, 6, 10, 12, 15, Table 2 and the relevant text, etc. Other detailed adjustments can be traced in the revised manuscript.

- We appreciate the suggested publication strategy of three papers. However, this work already is part of a larger strategy, with the field data interpretation and numerical modelling approaches being next. Therefore, we would prefer to focus more on the main goal instead of splitting it up.