A Summary of Revisions and Responses to the "Identification of watershed priority management areas based on landscape positions: An implementation using SWAT+" (Ref: HYDROL48272)

With regards to comments from the Editor:

"We have now received comments from three referees of your resubmitted article. While all three find the article to be of potential interest to Journal readers and a substantially revised version of the original manuscript, all three recommend additional major revisions before the manuscript can be considered for publication. Please carefully review the major suggestions provided by each of the three reviewers which include more clarification of definitions and methods employed and consideration of the improvement of model efficiency relative to accepted standards."

We appreciate the constructive suggestions from the three anonymous reviewers. We have carefully considered and responded to each comment. For the major concerns of the reviewers, we clarified the definition and information of landscape position units in three aspects, including the general concept, delineation method in SWAT^{+}, and parameter-settings for the study areas in the case study. The introduction of the classification method used in this study has been added. Regarding the improvement of model efficiency, we have carefully made some explanations and revision. We also created an open-source repository to store the modeling data and update the modeling details and results routinely according to the improvement of SWAT^{+} in the future. The link of the repository was provided in the revised manuscript. We hope the revised manuscript can solve the concerns of the editor and reviewers.

With regards to comments from Reviewer #1:

1. General comments:

"This study proposed adopting landscape positions along the hillslope as identification units of PMAs, so as to balance the general applicability and the representation degree of spatial heterogeneity. The logic of this paper is excellent, and I think the result can absolutely prove the advantage of using landscape position as identification units. However, the SWAT+ and Markov method are both not new so the authors should show what is new for this paper in a clearer way. Besides, I think this paper needs more evaluation data to support the modeling result, maybe it will more convincing."

Thanks for the reviewer's approval and comments. As the reviewer pointed out, the key innovation of this manuscript is proposing to use landscape position units within subbasins to identify PMAs to balance the general applicability to diverse geographic environments and the representation degree of spatial heterogeneity. The basic idea of implementing the proposed method is quantifying pollutants released from the landscape position units and distinguishing the pollutant load contribution of each unit to the watershed outlet (see the first paragraph of Section 2).

The exemplified implementation of this study adopted the $SWAT^+$ model to quantify the pollutants released and the Markov chain-based surrogate model to distinguish the source contribution. Specifically, the transition matrix of the Markov chain-based model was improved to represent both landscape position and channel units (see the second and third paragraphs of Section 2). Therefore, although the $SWAT^+$ and the Markov chain-based surrogate model are not new technologies, the combined application with corresponding improvements in demonstrating the proposed PMA identification method based on landscape position units is the first.

We have emphasized the main contribution and the new combined and improved method implementation in several places (including the first point of the highlight, the introduction of the exemplified implementation of the proposed method in Sections Abstract and Conclusion).

The evaluation data used to compare the PMA identification based on LSUs and subbasin units include the spatial distribution of PMAs and cumulative load contributions of spatial units. The spatial distribution qualitatively proved the finer and accurate identification of LSU-based PMAs than subbasins. The cumulative load contribution quantitively proved the better effectiveness of LSU-based PMAs than subbasins according to the relations between the area of PMAs and their total load contribution. We believe the two aspects of evaluation data are sufficient to prove the advantage of the proposed method. We hope this explanation is acceptable.

2. Specific comments:

1) "L29-L59 The disadvantages of "Subbasins, Artificial geographic and grid cells" have been introduced, but I think you should also list the main advantage of those identification units to compare with your idea."

In Section Introduction of the revised manuscript, we reviewed existing spatial units used in PMA identification in a progressive manner. For each type of spatial unit, the main advantage is introduced first, followed by the disadvantage. We have revised the third to fifth paragraphs of Section Introduction to clarify the advantage of each spatial unit. For example, "subbasin units are the most straightforward and frequently used identification units…the combination of subbasins as identification units according to administrative regions for the benefit of making and implementing watershed management policies, especially in large study areas", "artificial geographic entities have relatively homogeneous features from the perspectives of physical geographic processes and/or anthropogenic activities…are appropriate for use as identification units than subbasins in the corresponding geographic environment", and "raster cells are universal units to identify PMAs accurately using watershed models that explicitly represent flow routing among grid cells."

2) "L188-L190 The reason why using a retention coefficient capacity of pollutants to simplify the complicated channel routing process should be explained."

The complicated channel routing process of pollutants accounts for the chemical transformation or retardation of the substances. For example, a stepwise transformation from organic nitrogen to ammonia, then to nitrite, and finally to nitrate is simulated in SWAT/SWAT$^+$. For each channel of the study area, the difference between the output substance and the input can be explained by the retention effect of the channel, which is time-varying and affected by pollutant concentration, water temperature, and other factors. The yearly average retention of each channel can be regarded as its stable removal capacity of pollutants that can be calculated as a retention coefficient (Eq. 2 in the revised manuscript). The above explanation has been added in the second paragraph of Section 2.2.1.

3) "L227-229 We generally believed that inorganic nitrogen include ammonia nitrogen, nitrite nitrogen and nitrate nitrogen, and in most areas the ammonia nitrogen account for a large proportion. I think maybe you should explain why the nitrate nitrogen is only considered in inorganic nitrogen."

Thanks for the reviewer's comment. We agree that inorganic nitrogen generally consists of ammonia nitrogen, nitrite nitrogen, and nitrate nitrogen. We considered the organic nitrogen and nitrate nitrogen as the nitrogen load to identify PMAs based on LSU and subbasin units because of the modeling theory of the SWAT/SWAT$^+$. In SWAT/SWAT$^+$, the inorganic nitrogen output in the channel includes ammonia, nitrite, and nitrate nitrogen. The nitrate and organic nitrogen are relatively stable forms of nitrogen in the soil that are routed from HRUs into the channel with water and sediment (Neitsch et al., 2011). Since the nitrogen output at the LSU level is the sum of its internal HRUs' output, the nitrogen released from LSUs considered in this study also comprises nitrate and organic nitrogen. For the sake of simplicity, we would like to keep the use of total nitrogen (TN) to represent nitrate and organic nitrogen in the revised manuscript. At the same time, we added the above explanation in the last paragraph of Section 2.2.2.

4) "L241-248 What kind of classification methods are used in this study?"

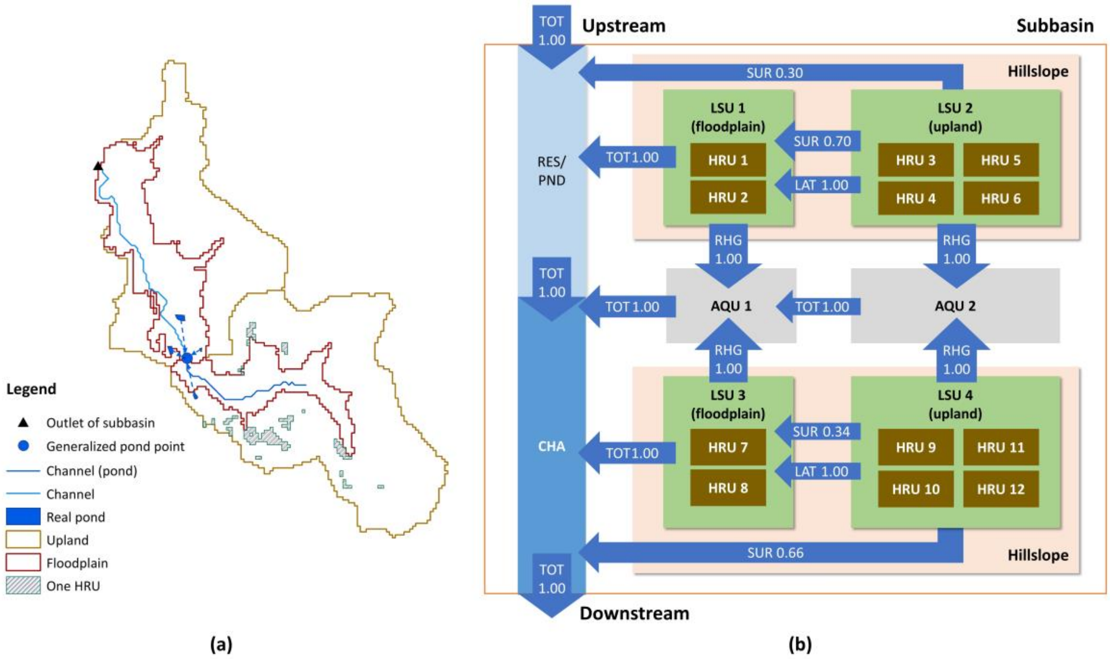
The natural breaks method was used as the classification method in this study. We have added the introduction of this method in the Section 2.2.3, i.e., "In this study, we adopted the natural breaks method, a commonly used classification method (De Smith et al., 2018; Giri et al., 2016), to classify the pollutant load contribution. The natural breaks method classifies the data into different classes with the statistical groupings and pattern characteristics inherent in the data to minimize the data difference within a class and maximize the difference between classes."

5) "L258 I think the information of landscape should be introduced."

Thanks for the reviewer's suggestion. We revised the introduction of the landscape position used in this study from three aspects. The first is the general concept introduced in the penultimate paragraph of the Introduction: "In this study, landscape positions refer to geographic objects that reflect the integrated effects of hillslope processes on topography and affect geographic processes on the surface. Landscape position units are universal in most geographic environments that can be delineated by slope position units."

The second aspect is the delineation method used in $SWAT^+$, as revised in the first paragraph of Section 2.1: "$SWAT^+$ uses the relative position index (RPI) of each cell in a gridded digital elevation model (DEM) to delineate LSUs. The RPI of each cell is the ratio of the drop length to its downstream valley (i.e., the stream cell) and the length from its upstream ridge cell to the same valley cell. The RPI ranges from 0 to 1. The cell with a RPI less than the user-specific threshold is classified as the floodplain." We also added a subfigure to Fig 2 (i.e., Fig 2a) to illustrate the spatial discretization scheme of $SWAT^+$.

The last aspect is the parameter-settings for the study areas, as revised in the first paragraph of Section 3.2: "The RPI thresholds for delineating uplands and floodplains are manually determined by visual interpretation of contour lines, which are 0.14 and 0.3 for the Zhongtianshe Watershed and the Willow River Watershed, respectively. In most situations, each subbasin has one upland and one floodplain. There may be an additional floodplain because of the very short channel generated after the setting of a pond or reservoir."



The revision of Fig. 2. Schematic of the spatial discretization scheme (a) and hydrologic connections between spatial units (b) implemented in SWAT+.

6) "L294-295 only 5-day monitoring data to calibrate total nitrogen the results if reliable?"

Thanks for the reviewer's comment. It is common in existing studies to manually monitor pollutants at several days' intervals for watersheds without automatic stations. In this study, the total nitrogen monitoring data at the Zhongtianshe watershed outlet has 181 values from 2014 to 2015. In the original manuscript, we use the 5-day monitoring data to generally describe the sampling interval of the observed data. In fact, the monitoring frequency was increased in the rainy season (i.e., from June to August), resulting in 53 values during the two years, which can be described as 3-day monitoring data. To make this data description clear, we revised the sentence to "The model performance for the total nitrogen was calibrated using the 5-day or 3-day monitoring data from 2014 to 2015 (a total of 181 values, of which 53 values during the rainy season were sampled in about three days interval from June to August in the two years), without validation." Therefore, we believe that the monitoring data of total nitrogen is reliable to calibrating the daily $SWAT^{+}$ model in our study. We also added the above description in Table 1 of the revised manuscript: "A total of 181 values were monitored in the Zhongtianshe watershed. During the rainy season (i.e., June to August), the sampling interval is about three days."

With regards to comments from Reviewer #2:

1. "The mathematics / modelling seem fine, but the key metric, the landscape position unit (LSU), isn't clearly defined. I think it is a metric that separates uplands and lowlands or specifically floodplains, but whatever it is, it doesn't have a clear definition and then description of how it was used to refine the subbasin scale."

Thanks for the reviewer's approval and comment. As stated in response to the first reviewer's fifth specific comment, we have revised the introduction of the landscape position used in the manuscript from three aspects. The first is the general concept in the penultimate paragraph of the Introduction. The second aspect is the delineation method used in $SWAT^+$ in the first paragraph of Section 2.1, i.e., using the relative position index (RPI) of each cell in a gridded digital elevation model (DEM) to delineate uplands and floodplains with a user-specific threshold. The last aspect is the RPI thresholds set for the study areas (the first paragraph of Section 3.2).

2. "Where are the N and P values across the spatial units to show how they are used as inputs?"

The N and P are released from each HRU with water and sediment in various states and lumped in the LSU, and then added to the channel (as shown in Fig. 2b and Eq. 9–13). Therefore, the pollutants released in the hillslope process can be obtained from the output of the LSU by the $SWAT^+$ model, which is the input data of pollutant sources for the Markov chain-based surrogate model (Eq. 5 in Section 2.2.2). The pollutants that flow in and out of each channel are used to calculate the retention coefficient for the Markov model (Eq. 2–3 in Section 2.2.1). In this study, the average annual total nitrogen (including nitrate and organic nitrogen; also see the response to the first reviewer's third specific comment) of the $SWAT^+$ model results during the calibration period were used as the input of the Markov model.

3. "There seems to be second objective which is an introduced methodology within the SWAT process – "improved Markov chain-based PMA identification method can be regarded as a method framework". If correct, then introducing two changes in a model (in this case scale and method herein) seems to undermine the attempt to demonstrate a change with the scale adjustment (LSU metric). For example, if something changes in the output, then is it the result of the method or the scale alteration?"

The proposed PMA identification method based on landscape position units was implemented by improving the original Markov chain-based surrogate model proposed by Grimvall and Stålnacke (1996). The improvement is to separate the pollutant release and retention process mixed in the subbasin, which depicts the pollutant release process on landscape position units and the retention in the channel unit, respectively (see the third paragraph of Section 2). The required input data of such improvement can be provided by the $SWAT^+$ model, i.e., the transition matrix of pollutants among LSU and channel units (see Section 2.2.1) and pollutants released on LSUs and pollutants that flow in and out of channels (see Section 2.2.2). Therefore, we claimed in the second paragraph of the Section Conclusions that the improved Markov chain-based PMA identification method can be regarded as a method framework to support more types of spatial units with explicit upstream-downstream relations and the corresponding watershed models.

So, the improvement in the proposed method does not change the calculation principle of the original Markov-based surrogate model. Thus, based on the same calibrated $SWAT^+$ model, the differences in identifying PMAs can be attributed to the identification units adopted (i.e., the LSU and the subbasin unit in this study). We added this sentence to the first paragraph of Section 3 to clarify the reviewer's concern.

4. "Figure 5 suggests there could be a further reduction in scale to HRU. This appears to 'improve' the results beyond LSU scale change. Why not present that HRU output and then discuss why the LSU is a better approach?"

Thanks for the reviewer's comment. The proper spatial units for PMA identification should follow three principles, as summarized in the literature review (see the 6th paragraph in the Introduction). The last but most important one is the spatial units should have hydrologic connections among each other. The HRU is delineated as the unique combination of soil, land use, and slope class within the LSU, which is spatially discrete (see Fig. 2a in the revised manuscript) and even lacks explicit spatial locations according to different delineation parameters. There are no hydrologic connections between HRUs in SWAT/$SWAT^+$. Therefore, HRUs are not suitable to be used as PMA identification units. We added this explanation in the first paragraph of Section 2.1.

5. "Regarding PMA designations – for the China watershed the difference is small and for the USA watershed it is much larger. This suggests that other/many watershed characteristics effect the model outputs, i.e., other factors may be overriding the impact of the chosen LSU metric. These types of uncertainty undermine your LSU conclusions (an example why many more watershed would help determine if the LSU is doing what you wish)."

To illustrate the effectiveness of PMAs identified on the LSU level compared to the subbasin level, the comparative experiment was conducted separately in two different study areas. The two case studies gave similar results in two perspectives, i.e., LSU-based PMAs improved the identification accuracy in spatial distribution compared with subbasin-based PMAs (see Section 4.1), and LSU-based PMAs contribute more load contribution than subbasin-PMAs on the same area (see Section 4.2). The results are sufficient to support the proposed idea of using landscape position units to identify PMAs.

We agree that many uncertainty factors affect the modeling results of the $SWAT^+$ model, which affects absolute differences in pollutant load contributions at the two levels in different study areas. However, this study is a “proof of concept” study (as pointed out by reviewer #3) to evaluate the value of LSU-based PMA identification, which means the comparison between different study areas is less important and beyond the scope of current study. From the two case studies with different geographic characteristics, we believe that the basic conclusion will keep the same, i.e., LSU-based PMAs are more accurate and suitable than subbasins in identifying PMAs for further watershed management. We have claimed this point in the last paragraph of Section 4.2: “Although absolute differences exist in the results of the two watersheds due to different geographic characteristics, the comparison between them is less important for the scope of this study (which is to evaluate the effectiveness of the PMAs at the LSU level). Instead, the similar appearance depicted by the relations between the area of PMAs and their total load contribution at the two levels in different watersheds can also show the universality and effectiveness of LSUs.” We hope this explanation is acceptable to the reviewer.

With regards to comments from Reviewer #3:

"…… In summary, I feel that the authors have done a reasonable job in trying to respond to the major comments of the original reviewers and have prepared a paper that may eventually be acceptable for publication in the Journal of Hydrology. The final acceptance of this paper for publication in the Journal of Hydrology depends on the authors revisions to this paper in response to the following Major and Minor comments. "

Thanks for the reviewer's approval and comments. We have carefully revised and responded to the comments. We hope the revision is acceptable to the reviewer.

1. Major Comments:

1) "Lines 303 and 304 state "calibrated SWAT+ models have approximately satisfactory performance proposed by Moriasi et al. (2007)", and, thus, the authors conclude on Lines 307 and 308 "both calibrated models are applicable for the validation of the proposed PMA identification method in this study." However, for satisfactory models Moriasi et al. (2007) indicate NSE should be greater than 0.5 and RSR should be less than 0.7, other than the flow validation for the Zhongtianshe Watershed none of the model results meet satisfactory levels. The PBIAS values for the Zhongtianshe Watershed are acceptable, but those for the Willow River Watershed are not. The nitrogen PBIAS value for calibration on the Willow River Watershed appears "good", but, in fact, a "good" nitrogen concentration result with a poor hydrologic simulation yields poor nitrogen load prediction. Thus, in my experience with dozens of watershed models I would not categorize either SWAT+ model as satisfactorily simulating flow or total nitrogen. The results of the current study would be more convincing if the SWAT+ models were actually satisfactory. For example, since the Markov retention coefficients are derived from the internal loading information from the SWAT+ model, the current coefficients, derived from unsatisfactory models, are unreliable.

On the other hand, the current paper is not designed to guide watershed managers in the design and implementation programs for best management practices in the study watershed, but rather this is a "proof of concept" study on the value of landscape unit scale PMA determination. Thus, truly reliable modeling results may not be necessary. However, I think the paper would be more convincing if the authors could improve the calibration and validation of the SWAT+ models, and rederive the PMAs."

Thanks for the reviewer's comments. We agree that the performances of daily model built for the two study areas in this study did not reach the satisfactory standards based on the evaluation criteria for monthly models as proposed by Moriasi et al. (2007). The daily model is more likely to have worse performance according to such satisfactory standards. We have tried to improve the model performance in the last several months but still cannot fit such a satisfactory standard for monthly models. We created an open-source repository to store the modeling data and would like to update the modeling details and results routinely in our following study according to the active improvement of SWAT+ (https://github.com/lreis2415/WatershedModelingData). Besides, considering this study mainly utilizes the relative rather than absolute reliable model results to verify the effectiveness of the proposed PMA identification method, we believe that both the calibrated models can be regarded as acceptable for this study. We have revised the last paragraph of Section 3.2 accordingly to clarify this point. We hope the revision is acceptable to the reviewer.

2) "Lines 315 and 316 state that the "natural break method was utilized to classify the nitrogen load contribution of the spatial units". The paper should include details on this method and cite appropriate references when giving these details."

We have added the details and corresponding references of the natural break method in the Section 2.2.3 of the revised manuscript, i.e., "In this study, we adopted the natural breaks method, a commonly used classification method (De Smith et al., 2018; Giri et al., 2016), to classify the pollutant load contribution. The natural breaks method classifies the data into different classes with the statistical groupings and pattern characteristics inherent in the data to minimize the data difference within a class and maximize the difference between classes."

2. Minor Comments:

1) "Lines 51 and 52 state (note, text slightly revised): "they are not easily generalized as generally applicable identification units and are widely applied." This statement seems contradictory, that is, if artificial geographic entities are not easily generalized as applicable identification units, then why are they widely applied? Please re-write and clarify."

We revised this sentence to state our opinion more clearly, i.e., "Although artificial geographic entities are appropriate for use as identification units than subbasins in the corresponding geographic environments, they are not easily generalized as universal identification units and, thus, cannot be widely applied to diverse geographic environments."

2) "Numerous editorial suggestions are made throughout the marked manuscript which the authors should consider when preparing a revised version of this paper for re-review."

We deeply appreciate the careful review and helpful editorial suggestions. We adopted most suggestions according to the marked manuscript and retained a few for keeping the precise meaning (as list below, marked in bold and underlined):

a. The fourth point in the Highlight: "LSU-based PMAs are more effective in distribution **<u>and</u>** cumulative load contribution." – The "distribution" means actually "spatial distribution."

– We used “distribution” to meet the limited number of characters of the highlight. Therefore, the “and” was not changed to “of” as the reviewer suggested.

b. Line 9 and 10 in the Abstract: “Existing spatial units used to identify PMAs are commonly based on three **concepts** including subbasins, artificial geographic entities, and grid cells.” And Lines 24-27 in the Introduction: “The identification units utilized in existing research are mainly based on three **concepts**: subbasins (Shang et al., 2012; Chen et al., 2014; Shen et al., 2015; Dong et al., 2018), artificial geographic entities (Tian et al., 2020), and grid cells (Kovacs et al., 2012).”

   – In our opinion, these three types of spatial units are not differentiated by apparent spatial scales. For example, the spatial scale or granularity of artificial geographic entities depends on specific applications which may be similar to the subbasin scale. Therefore, we used the “concepts” here instead of the “spatial scales” suggested by the reviewer.