Response to comments by Referee #1

We thank Anonymous Referee #1 for providing useful and constructive comments. We will carefully revise the manuscript and address all the points raised by the referee, as this will clearly improve the quality of our manuscript. Particularly, we intend to replace the linear regression model that was criticized by Referee #2 by a qualitative analysis.

The authors present an interesting study design with almost experimental character. Based on Fig. 1, one can conclude that there is a spatial separation between different crops in the catchment (arable crops versus vineyards). Because these crops receive different pesticide treatments (with regard to timing, compounds) one gets a signal in the outlet of the catchment that is related to a certain spatial unit. This design has the potential for learning how spatial aspects, physico-chemical properties of pesticides and agricultural practices influence pesticide losses (dynamics, loss rates). This could also be relevant for better understand the functioning of the wetland as is the main purpose of this manuscript.

Reply: We thank referee #1 for acknowledging the unique character of our experimental setup to investigate different types of contaminant input signals and their mitigation in an agricultural wetland. We will highlight these aspects more clearly in the updated version of our manuscript.

The manuscript analyses the effect of a wetland at the outlet of this agricultural catchment on pesticide transfer further downstream. Based on samples taken during 10 events and five baseflow periods, the retention capacity is studied for four active ingredients and two metabolites regarding the relative reduction of peak concentrations and pesticide mass transported downstream. Specifically, the authors investigated the effect of the shape of the chemograph at the inlet and a few compound properties and explain the differences in the chemograph shapes of different compounds by being applied to different parts of the catchment.

Monitoring pesticide dynamics in natural systems is demanding and requires substantial efforts in the lab and in the field. The observed pesticide behaviour in the environment is complex and often influenced by specific conditions at the local scale. Therefore it is valuable for science and practice if such observations get published from different locations and catchments because it broadens our understanding of the environmental fate of pesticides under different conditions.

Unfortunately, there a number of issues that limit the value of the manuscript in its current status. I describe my main concerns below and add some further details at the end.

Scientific quality:

The data analysis is rather superficial and several of important conclusions are not strongly backed up with data. This holds true for example for the interpretation of the different dynamic patterns identified (L. 210 - 221). For example, the authors hypothesise that fungicides from the upstream vineyards are more quickly transported to the stream than the herbicides applied closer to the creek in the valley bottom (L. 210 - 213). Why this should be the case remains unclear. Checking a previous publication on the catchments (Gassmann et al., 2012) reveals that there is a dense road network connected to the pipe drains. These structures are made explicit in the earlier publication but no linear structures are indicated in the catchment map of this manuscript. Given the fact that pesticide drift to non-target surfaces such as roads may be important for pesticide transport in vineyards (Lefrancq et al., 2013) important aspects of the catchment are neglected and not included into the discussion of the results.

Reply: We agree that structures like roads may represent shortcuts and accelerate pesticide transport. Particularly spray drift is an important process here, which is more relevant for fungicides, as they are applied into the foliage and not close to the ground as herbicides. We are thankful for this hint and will discuss this in detail in the revised manuscript.

Also temporal aspects are only treated superficially. The shape of the chemograph of a pesticide is strongly influenced by the rainfall dynamics. Because the authors compare the dynamic of different pesticides across different events (this is not really evident from the main text, but see the SI), differences in concentration dynamics could also be strongly influenced by rainfall patterns and discharge behaviour. Unfortunately, no respective data is shown or discussed.

Reply: Description of the experimental setup and the included data will be revised to make it clearer. A figure containing concentrations and discharge during all events will be provided and effects on chemographs will be discussed.

This holds also true for the timing between the last application of a pesticide and the rainfall event. The authors do not discuss this aspect and treat all compounds and all events the same (except for two events for flufenacet where too many samples $<$ LOQ). However, inspection of the actual concentration data reveal strong differences in the concentration levels of the different compounds across all events. High concentrations of several hundred to thousands of ng L^{-1} were found for boscalid for all events, while flufenacet was only found in one event above 1600 ng L^{-1} but otherwise never above 40 ng L^{-1} . Obviously, the history of the compounds since the last application was very different. Neglecting such aspects but interpreting the different relative concentration dynamics with respect to transport differences from different parts of the catchments is not very solid.

This also limits the value of the cluster analysis. The results in Fig. 3 show that a given compound appears in different clusters. However, this aspect is not properly discussed and no explanation is provided why this was the case nor what this implies for interpretation of these clusters.

Reply: We fully agree that the chemograph shape is influenced by many factors including compound properties, source areas, transport pathways, absolute concentrations, rainfall and discharge dynamics and also pesticide application time, some of which we have not addressed to the necessary level of detail so far. In general, we see three groups of factors that may influence chemographs. These are compound properties (e.g. sorption affinity, degradability), event properties (rainfall and discharge dynamics) and catchment properties (application areas, rates and timing, transport pathways, slopes). In the investigated catchment, different pesticide types are applied to different compartments of the catchment so that differentiation between compound and catchment-related factors is challenging. However, we consider our analysis valuable as it revealed that (1) systematic differences were evident in chemographs and (2) clustering was mainly according to compound types, rather than according to events. This suggests that the compound type (and co-occurring catchment-related factors, e.g. source areas and transport pathways, application rate and time) had larger influence on chemograph shapes than event properties. We are thankful to the reviewer, since obviously we did not make these points clear enough and will discuss them in more detail in a revised manuscript.

Additionally, it is not clear to which degree these results reflect the full spectrum of observed pesticide dynamics. The authors mention in the Method section (L. 190) that they have removed outliers based on purely statistical analyses. I don't think the procedure is sound (see below) and may bias the findings by excluding unusual - or simply rare - dynamics.

Reply: No outliers were removed for the cluster analysis so that this issue is limited to the regression model. As we intend to replace the linear regression model by a qualitative data analysis in response to objections by referee #2, we consider this point obsolete.

Finally, the calculation retention rates raises a number of question marks, which may further impact the subsequent linear model for describing the retention efficiency of the wetland. First, given the measured concentrations (see SI, but also Fig. 3) it is evident that in many cases the last data point does not reflect baseflow concentrations after the event. Accordingly, the mass loss during the events may have been substantially larger in some cases. The extent of this effect depends on the unobserved concentration dynamics but also on the discharge. Unfortunately, no discharge data is provided that illustrate which part of the event hydrographs have actually be covered by the sampling.

Reply: We are aware of the issue with elevated concentrations in the last event sample and calculated mass flux during the events in order to assess the effect on mass balances (R_M). Despite elevated concentrations, mass flux was close to zero at the end of sampling in most cases. We agree that this information should be accessible to the reader and will provide a figure showing mass fluxes in the supplementary material of a revised manuscript. Late time concentration was largely irrelevant for R_C as peaks normally occurred in earlier samples. R_C may, however, be influenced by discharge dynamics. A figure for this will be provided in the revised manuscript so that the reader obtains an idea on associated uncertainties.

Second, the observed concentration levels demonstrate that for compounds such as flufenacet most events reflect conditions long after the last application. Accordingly, the concentration signal is rather weak and calculated retention is laced with considerable uncertainty. This aspect is not mentioned.

Reply: We agree that relative uncertainty is higher at lower concentration and will include this aspect into the discussion of a revised manuscript.

Scientific significance:

The issues listed above reduce the scientific significance of the manuscript. Additionally, there are questions about the scientific insight conveyed by the manuscript. There are two major issues:

Relevance of the shape of the chemograph The effect of the chemograph is almost a trivial finding. Given the short residence time in the wetland (about 1 h; see L. 81) degradation processes will be very limited and the main effect on peak concentrations in the outlet is expected by dispersion (as expected by the authors, see below), i.e. mixing water of different concentrations at the inlet. Obviously, the more variable the input, the larger the relative effect of mixing on the relative maximum concentrations. Actually, the authors have stated this outcome already in the Introduction as fact: "'*Peak concentration reduction will be stronger for a signal with pronounced peak and low background than for a signal with small peak and high background if both signals are exposed to the same dispersion.*"' (L. 54 - 55).

[... For mathematical derivation by referee #1 see original comment …]

Reply: We agree that the finding that different signals are affected differently by dispersion is not novel and we did not claim that it is, although we are not aware of many wetland studies that explicitly address this aspect. Instead, contaminant peak reduction is often shown as an important mitigation effect of wetlands, as it decreases acute toxicity (Bundschuh et al., 2016; Elsaesser et al., 2011). However, this is only one aspect of the problem, because peak reduction does not necessarily mean that contaminant mass is also reduced. In our study (sampling scheme, data analysis, etc.) we are addressing both types of toxicity (acute and permanent). We are aware that this aspect was not communicated clearly enough in our manuscript and we will revise our updated version correspondingly.

In this context I am furthermore surprised that the authors did not use the model that was developed for the study wetland for evaluating the effect of different chemograph shapes (Schuetz et al., 2012). Although the wetland may have undergone changes since the last tracer experiments, it would be a useful null model for testing how different input signals influence the reduction in peak concentrations.

Reply: We thank referee #1 for this suggestion and agree that a proper transport model would be useful. In fact, we experimented quite a lot with possible representations of the investigated system in the solute transport model (OTIS) used by Schuetz et al. (2012). Although we were actually able to reproduce concentrations at the basin outlet from those at the inlet acceptably well, our confidence in the model was low and we decided to not include the model for the following reasons:

(1) For evaluation of different input signals, it is crucial that solute transport (incl. dispersion) was well parameterized. OTIS was designed for stationary flow-conditions. Application of OTIS during transient flow is technically possible. However, we consider the parameterization in such cases questionable as model parameters that are obviously time-variant have to be assumed constant, e.g. storage zone area,

dispersion coefficient, and exchange rate. Although a conservative tracer injection was performed during one of the discharge events, we were unable to identify a range of transport parameters that was plausible when compared to the discharge conditions during the events.

(2) Parameterization of transport clearly influenced process-related parameters, i.e. rates of decay and kinetic sorption, so that the latter could hardly be estimated from the model.

(3) Comparison with Schuetz et al. (2012) was not possible because their model was based on stationary discharge, while we were dealing with event data and the studied system was fundamentally changed by the implementation of the retention pond in 2016, between the experiments by Schuetz et al. (2012) and the start of sampling for the present study.

Therefore we came to the conclusion that the use of an OTIS-type model and the interpretation thereof rather introduced additional uncertainties and that further insights provided by the model were limited.

How to generalise the findings? It is difficult to generalise any of the findings reported in the manuscript because they are very context dependent and results such as the cluster analysis are very phenomenological. The aspect that the shape of the chemograph has an influence on the relative degree by which peak concentrations are reduced is quite evident for the type of system under investigation.

Reply: We agree that findings from field experiments often depend on local conditions. However, regarding processes in the catchment, we do not consider the results of the cluster analysis “very phenomenological”. We showed that the chemograph shape more strongly depended on catchment or compound properties and not on event characteristics. Potentially influential factors that can be separated by this type of analysis may be different in other catchments, e.g. source areas and compound properties might be distinguished more easily if application areas of the compound groups were not spatially separated as in this study. We therefore see a high potential for this type of analysis in other catchments, although local conditions have to be considered. We thank referee #1 for this objection as it shows that we did not communicate this aspect clearly enough. We will revise our manuscript accordingly.

We intentionally focus on the chemograph shape as it links processes in catchments to those in treatment wetlands and we consider this aspect in principle transferable to other systems. Although the a larger dispersion of sharp peaks is not novel, the importance of this relationship is generally not reflected in existing literature of contaminant mitigation so that we consider our work a valuable contribution to the body of knowledge in this field.

Detailed comments:

L. 30: The phrase "‘... which may be equally or more mobile, persistent and toxic than their PC ..."’ is misleading because it does not mention the general case that transformation products are less toxic.

Reply: This will be stated more clearly in a revised manuscript.

L. 93: How often were grab samples taken?

Reply: Grab samples were taken at 7 occasions, after careful inspection of the hydrograph, we conclude that 5 of the sampling in fact represented stationary flow conditions. This will be stated more clearly in a revised manuscript.

L. 123: How adequate is it to only take one single isotope-labelled internal standard not corresponding to the target compounds? Checking these compounds in one of our current analytical methods, retention times vary between 16.7 (metazachlor-ESA) and 21.0 min (penconazole). Also the KOC values vary by a factor of 400 between these two compounds. Please provide additional information supporting the assumption that terbutryn as an adequate internal standard (e.g., recoveries).

Reply: We will provide information accordingly in the manuscript.

L. 129 - 130: Please check the correct number of significant digits (can you measure with a precision of tens of picograms per litre?).

Reply: We will re-check the precision of our measurements and adopt the number of digits.

L. 142: Please provide the version of R. I assume that you did not implement the algorithm but used `kmeans()` implemented in standard R.

Reply: The version of R will be provided.

L. 174 - 176: Why did you include DT_{50} a priori? I'd recommend to leave it in. The subsequent analysis would reveal whether or not is had any statistical relevance.

Reply: As we intend to replace the linear regression model by a qualitative data analysis in response to objections by referee #2, we consider this point obsolete.

L. 179: How did you quantify the water balance error? Please explain.

Reply: What we called the water balance error is the relative difference between water masses registered at both gauges G1 and G2. We acknowledge that this term may be misleading as it does not necessarily represent a measurement error. We will call this parameter "Relative water balance anomaly" in a revised manuscript and describe how it was calculated.

L. 190: The definition of outliers and their handling is not sound. Cook's distance is simply a mean of identifying data points that deviate in a statistical sense from the rest of the data population and have a strong influence on a derived regression model. This does not imply that the data point corresponds to an outlier that can be discarded from the analysis! It may be that the outlier reflects reality as well as all other data. They may simply reflect rare events. Of course it is important to assess the influence of statistical outliers on model performance and predictions. However, unless there are sound reasons to exclude data as outliers because these reasons indicate the outliers to be wrong, outliers have to be included in the analysis. For example, it can be made transparent that some data (explicitly shown) deviate from the others in a specific way and discuss possible reasons. Hiding them to the scientific community introduces bias.

Reply: We acknowledge that we should have been more transparent about handling of outliers in the manuscript and will carefully address this point in a revised manuscript. Outliers were only excluded in the regression model but not in the cluster analysis.

References

Bundschuh, M., Elsaesser, D., Stang, C., and Schulz, R.: Mitigation of fungicide pollution in detention ponds and vegetated ditches within a vine-growing area in Germany, Ecol. Eng., 89, 121–130, doi:10.1016/j.ecoleng.2015.12.015, 2016.

Elsaesser, D., Blankenberg, A.-G. B., Geist, A., Mæhlum, T., and Schulz, R.: Assessing the influence of vegetation on reduction of pesticide concentration in experimental surface flow constructed wetlands: Application of the toxic units approach, Ecol. Eng., 37, 955–962, doi:10.1016/j.ecoleng.2011.02.003, 2011.