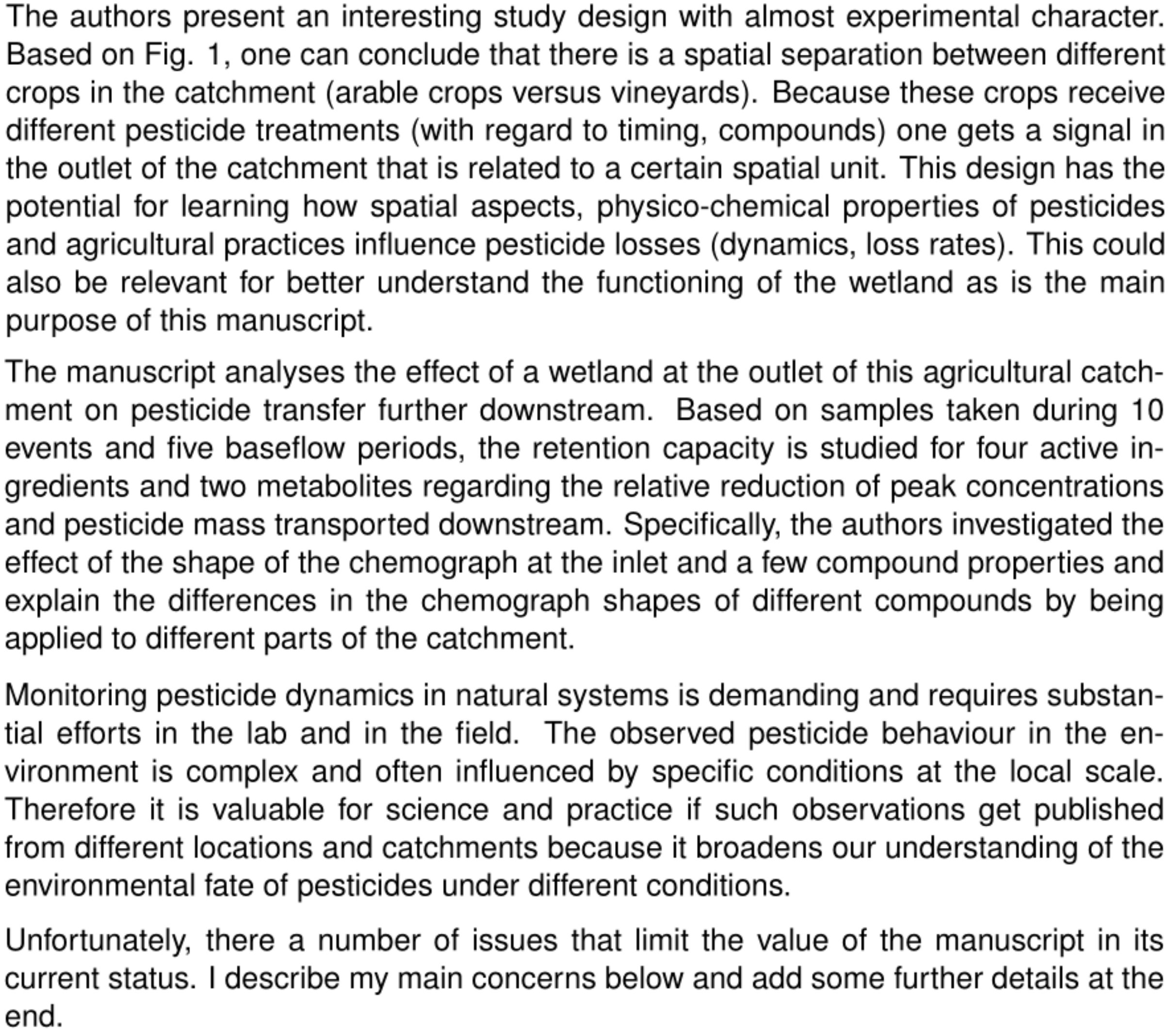
Dear Dr. Zehe,

We are thankful for the constructive comments by the editor and both referees which helped us to considerable improve the quality of our manuscript. Below please find our revised manuscript with point-by-point answers to the comments of the referees. In the document below referee comments are represented in black and our responses in blue. Italic font indicates quotations from the revised manuscript. Line numbers refer to the mark-up version of the manuscript below.

Response to comments by Referee #1



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.

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:

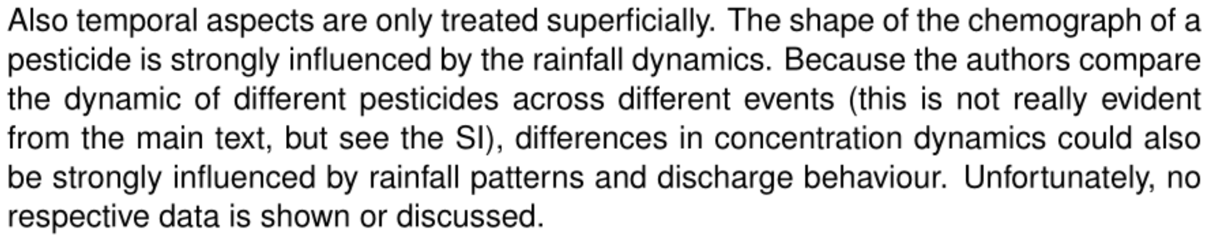
The data analysis is rather superficial and several of important conclusions are not strongly backed up with data. This holds true for example for the interpretation of the different dynamic patterns identified (L. 210 - 221). For example, the authors hypothesise that fungicides from the upstream vineyards 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. We improved the description of the catchment accordingly, added the primary drainage network to the map of the catchment (Figure 1), and discussed different mobilization dynamics in more detail. The description of the now reads as follows:

> *"The study site (Figure 1) is located inside a flood detention basin in the 1.8 km² Loechernbach catchment, southwest Germany. Catchment elevation ranges between 213 m.a.s.l. at the outlet and 378 m.a.s.l. in the western corner. Mean precipitation was 800 mm a^{-1} between 2009 and 2018, mean air temperature 11.3 °C. Soils mainly consist of calcaric regosols which formed on aeolian loess and have a typical grain size distribution of 10 % sand, 80 % silt and 10 % clay. Most of the catchment is dedicated to large artificial vineyard terraces (71 %), while croplands occupy the valley bottoms (20 %). Forest only accounts for a small portion (9 %) and is limited to the most elevated part of the catchment. This partition in land use is reflected in the main application areas of pesticide types. Fungicides are applied on vineyard terraces, while herbicides are mainly applied to the cropland in the flat valleys. Large parts of the catchment are drained by a sub-surface pipe network (Figure 1) connecting vineyards and paved roads to the main channel in the valley. This drainage network causes fast downstream transport of storm water and suspended sediments (Gassmann et al., 2012). In addition, fields in the valley bottoms are drained by a secondary network of smaller and usually shorter field drains that are either connected to the primary drainage network or directly connected to the stream (Schuetz et al., 2016). A 20,000 m^3 detention basin was built at the outlet of the Loechernbach to prevent flooding of the downstream village."* (Lines 98-112)

More emphasis is now laid on hydrological short cuts in the interpretation of the chemograph clusters, particularly regarding cluster B:

> ***"****The fact that concentration in cluster B quickly increased with discharge (within 30 minutes) is in line with fast transport from the vineyard terraces to the stream via roads and drainage pipes as described by Gassmann et al. (2012) for suspended solids in the studied catchment. Along such preferential pathways, compound properties, such as sorption affinity, may be less important (Gomides Freitas et al., 2008) compared to e.g. percolation through the soil with intense contact to sorption sites in the soil matrix. Moreover, fungicides are applied by sprayers into the foliage and can drift to e.g. paved surfaces from which they can be quickly mobilized by subsequent rainfall (Lefrancq et al., 2013). We therefore hypothesize that cluster B was mainly produced by surface flushing and fast transport pathways of fungicides. This explained the quick rise and subsequent decline in concentrations (concurrent with plateaus produced by slower flow components in cluster A).****"*** (Lines 574-583)



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: We improved the description of the overall aim of this study:

> *"We hypothesize that the efficiency of contaminant mitigation in VTSs depends on the shape of the input chemographs and eventually on the factors that produce these signals in the catchment. In order to test this hypothesis we grouped chemographs of 6 contaminants mobilized in a viticultural catchment during 10 discharge events into clusters according to chemograph shape. We then compared peak concentration reduction and mass removal in a downstream VTS both among clusters and in terms of compound properties and discharge dynamics."* (Lines 74-78)

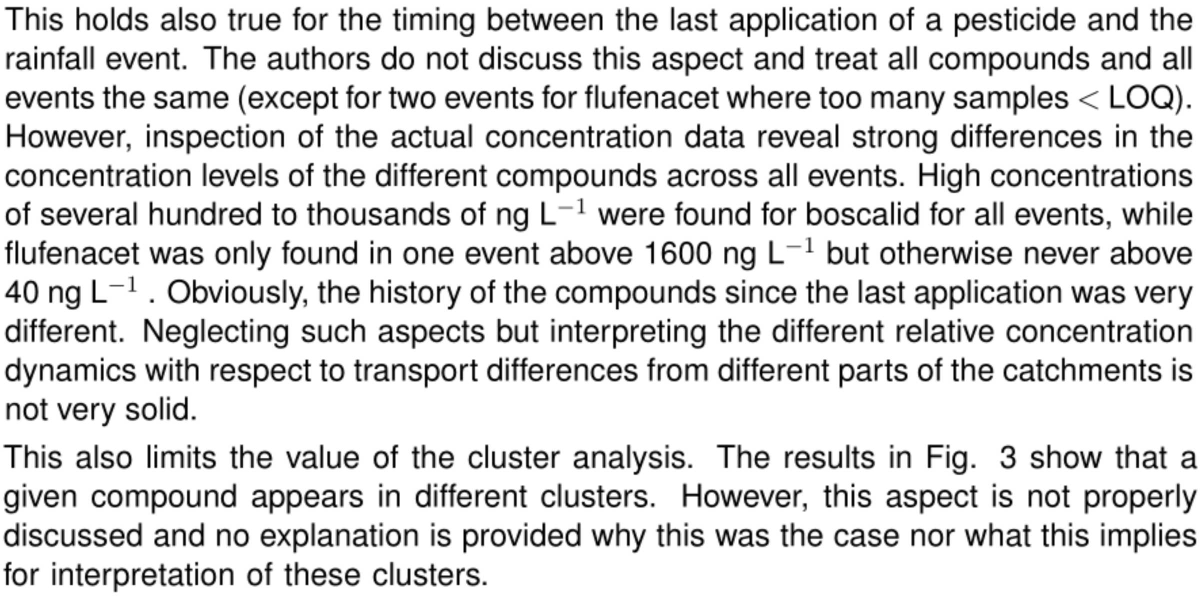
Different discharge dynamics of the events are now described qualitatively:

> *"The 10 events were characterized by different discharge magnitudes and dynamics (Figure 3). Mean discharge during the events ranged between 0.7 (E10) and 32.0 L s*$^{-1}$ *(E2) with respective peak values between 4.4 (E10) and 199.7 L s*$^{-1}$ *(E2). The recorded event hydrographs included events with one single discharge peak (E4, E5, E6, E10), with one major peak followed by one or more secondary peaks (E2, E3, E7, E9), and events in which a major peak followed an earlier smaller peak (E1, E8). In most cases discharge had recessed to pre-event levels by the end of the 12-hour sampling procedure, only E1 and E2 showed ongoing flow recession."* (Lines 330-337)

The influence of discharge dynamics on chemograph shapes was added to the discussion of chemograph clusters, particularly of cluster D:

> *"Cluster D included chemographs of both herbicides and TPs and presented a clear peak that was often defined by a single sample 2 h after the beginning of the event. In contrast to cluster A, cluster D was characterized by a single sharp discharge peak (except in event E7 where a second peak occurred shortly after the first) and mainly included chemographs during periods of low flow. Our interpretation is that cluster D represented flow events in which no dilution of herbicide and TP fluxes by fungicide fluxes or secondary discharge peaks occurred. Low pre-event discharge in cluster D compared to cluster A may indicate low water levels which may have caused a slower response as no enriched pre-event water was released from the soils in the valleys."* (Lines 594-600)

In order to provide a better idea of our data set, we replaced Figure 2 in the original manuscript, by a new figure (Figure 3) showing discharge dynamics and pesticide concentrations during all 10 events both at the inlet and the outlet of the detention system.



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.

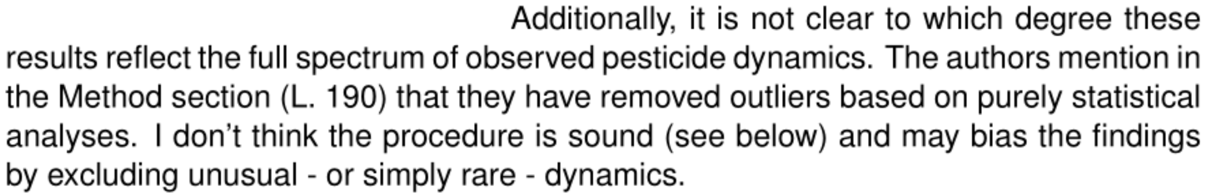
This also limits the value of the cluster analysis. The results in Fig. 3 show that a given compound appears in different clusters. However, this aspect is not properly discussed and no explanation is provided why this was the case nor what this implies for interpretation of these clusters.

Reply: We fully agree that the chemograph shape can be influenced by many factors and our analysis is not able to account for all of them. We do not have data on exact pesticide application times and rates, but included this aspect in the discussion. Given the spatial separation of fungicide and herbicide application areas, it seems highly plausible that these groups of compounds show different transport dynamics. In the revised manuscript we discussed the chemograph clusters and factors that cause their differentiation in more detail. The potential effect of recent application was particularly addressed with respect to cluster C:

> *"Cluster C was composed of chemographs of all compounds but mainly from events E2 and E4 indicating event dependence. Two aspects were found to support this idea. First, there was a secondary discharge peak in event 2 that did not contribute much in terms of contaminant concentration but rather caused dilution and produced particularly flat chemograph tails. Second, peaks of herbicides and TPs were less delayed compared to fungicides. This may be the result of recent herbicide application and active surface runoff in the flat valleys. Timing of pesticide application was identified as the main export driver of currently used pesticides by Imfeld et al. (2020) who performed a cluster analysis on rainfall data from a headwater vineyard catchment. Based on the magnitude of discharge and amount of mobilized contaminants (concentration of metazachlor* $\approx 10\ \mu g\ L^{-1}$*), both explanations seem plausible in event E2. Event E4, however, did not show particularly high herbicide concentration nor a secondary discharge peak. Although it is obvious that chemograph shapes in cluster C differed from the other clusters, unfortunately, the responsible factors remain unclear."* (Lines 584-593)

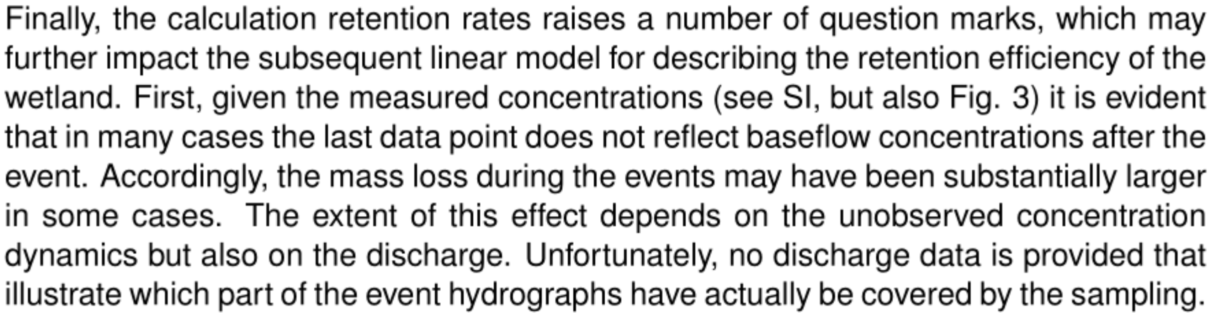
We also added a paragraph addressing the effect of overlooking important aspects on the interpretation of the clusters:

> *"The unclear interpretation of cluster C suggest that we missed important factors for the formation of chemographs. In fact, variables like spatial distribution of rainfall or pesticide application rates and timing (Imfeld et al., 2020) and possibly other factors likely influenced chemograph shapes. Knowing all these variables would not change the results produced by the clustering algorithm but rather increase our ability to interpret them. Nevertheless, the cluster analysis helped to explore how the catchment and processes therein influenced concentration signals of mobilized contaminants."* (Lines 601-608)



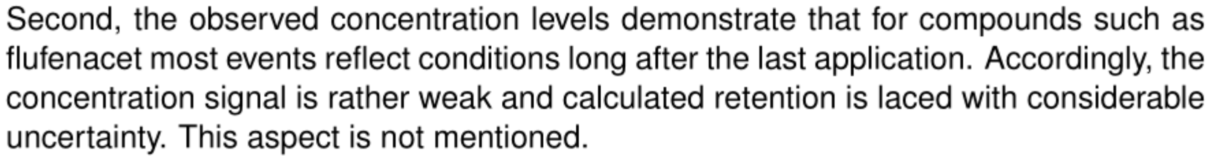
Additionally, it is not clear to which degree these results reflect the full spectrum of observed pesticide dynamics. The authors mention in the Method section (L. 190) that they have removed outliers based on purely statistical analyses. I don't think the procedure is sound (see below) and may bias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

Reply: Following referee #2 (see below) we excluded the regression model for which we had removed outliers. In the present analysis no outliers were removed.



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: see below.

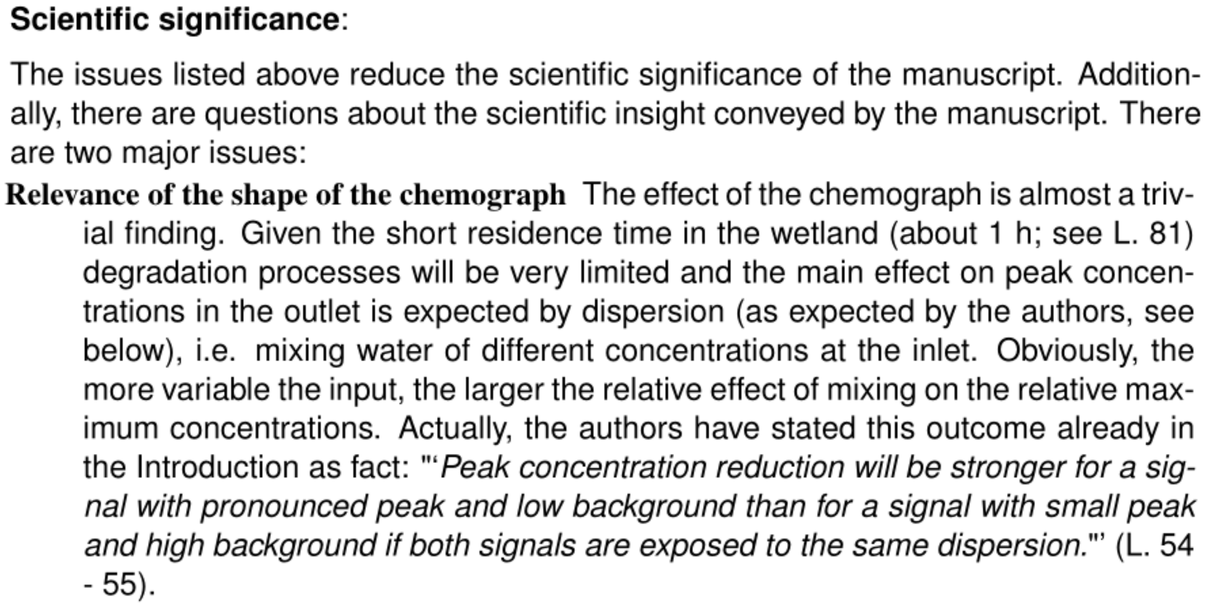


Second, the observed concentration levels demonstrate that for compounds such as flufenacet most events reflect conditions long after the last application. Accordingly, the concentration signal is rather weak and calculated retention is laced with considerable uncertainty. This aspect is not mentioned.

Reply: We added a section to the discussion dealing with uncertainties of the monitoring setup and added a figure to the supplementary material showing that contaminant mass flux was usually very low at the end of the sampling procedure (Figure S1):

> *"Regarding chemographs and calculation of R_C, uncertainties arose from timing and frequency of sampling and analytical error, and additionally from discharge measurement when calculating masses and R_M. Analytical methods used in this study usually produced very consistent results so that variability in concentrations of parent compounds in duplicate samples was low (sd < 10 %). However, in individual samples collected at G1 analytical variability was elevated for met-ESA and met-OA (Figure 3), reducing confidence in concentrations and the derived measures R_C and R_M of TPs in the affected chemographs (E2, E5, E7, E8, E9). Uncertainty related to timing and frequency of sampling can hardly be quantified but certainly depends on how well the sampling intervals captured variability in concentrations during flood events and how well the time lag between upstream and downstream sampling matched the residence time of solutes in the wetland. Lefrancq et al. (2017) assessed the effect of sampling frequency in pesticide monitoring data collected during runoff from a single vineyard and found that acute toxicity of pesticide flushes was underestimated up to 4-times when calculated from event means and up to 30-times when calculated from random samples. Although these data were collected on the plot scale and we assume that variability in our catchment is lower due to longer flow paths and mixing processes on the catchment scale, uncertainty of the chemographs in our study could have been reduced by increasing sampling frequency. Regarding the timing of upstream and downstream sampling, there is evidence that water residence time in the wetland was in fact shorter than one hour. The observation that for quickly responding compounds, such as boscalid, concentration in the first sample at G2 was often elevated compared to the first sample at G1 indicates that the contaminant flush hat already reached G2 when sampling started. This did not influence determination of*

$C_{out,max}$ and R_C in the outlet of the wetland, as concentrations were still rising from the first to the second sample (Figure 3). However, effects on M_{out} were higher, since a relevant fraction of contaminant mass leaving the wetland was not registered and thereby caused overestimation of R_M (Figure S1). Another source of uncertainty exclusively affecting contaminant mass and not concentrations was the use of different gauging systems at G1 and G2. Different shapes of the measurement cross-section (triangular at G1 and rectangular at G2) caused G2 to be less precise and water imbalances on the event scale, particularly when flow was low. Summarizing the setup constraints above, we have high confidence that the experimental setup produced realistic chemograph shapes and captured peak concentration reasonably well, but are less confident regarding contaminant loads." (Lines 524-547)



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:

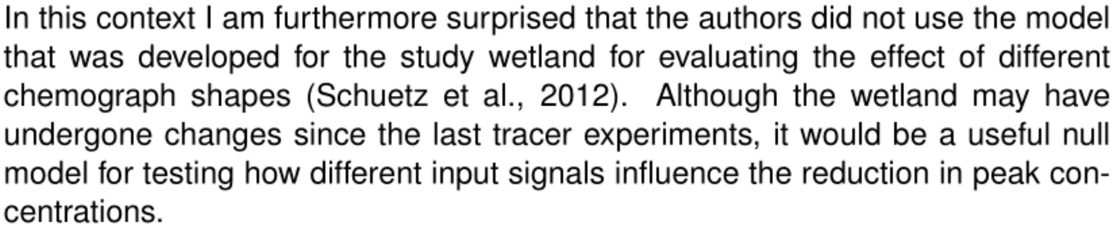
Relevance of the shape of the chemograph The effect of the chemograph is almost a trivial finding. Given the short residence time in the wetland (about 1 h; see L. 81) degradation processes will be very limited and the main effect on peak concentrations in the outlet is expected by dispersion (as expected 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 do not claim it to be so. However, the shape of the input chemograph is usually not considered in studies on contaminant mitigation in studies. Our data show that signal shape is an important explanatory variable when assessing the functionality of VTSs, particularly when residence time is low and most of the mitigation in VTSs is due to dispersion. We made this point clearer by revising the definition of the research gap in the introduction:

"Regardless of whether VTSs target concentration reduction or mass removal, mitigation efficiency is usually associated with physicochemical properties of target compounds (Vymazal and Březinová, 2015) or VTSs, including their operation mode (Tournebize et al., 2017). However, following the concept of advective-dispersive transport (Fischer et al., 1979), the mitigating effect

> *of dispersion on a concentration signal does not only depend on the magnitude of dispersion but also on the shape of the signal. Peak concentration reduction will be stronger for a signal with a pronounced peak and low background than for a signal with a small peak and high background if both signals are exposed to the same dispersion. Chemograph shapes, in turn, are dictated by processes in the contributing catchments. The influence of this chain of effects on contaminant mitigation and hence VTS efficiency has not been systematically investigated so far.”* (Lines 65-73)



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.

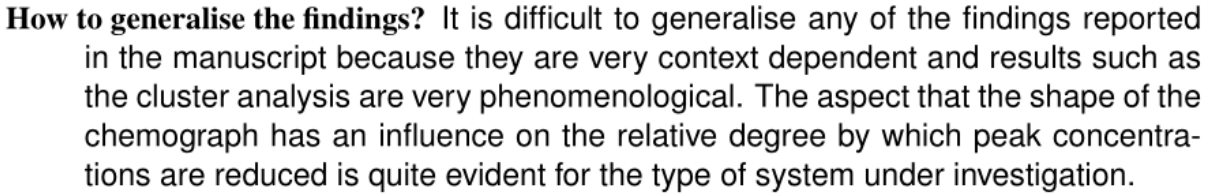
Reply: We thank referee #1 for this suggestion and agree that a proper transport model would be useful. In fact, we experimented quite a lot with possible representations of the investigated system in the solute transport model (OTIS) used by Schuetz et al. (2012). Although we were actually able to reproduce concentrations at the basin outlet from those at the inlet acceptably well, 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 adequately parameterized. OTIS was designed for stationary flow-conditions and as such used by Schuetz et al. (2012). 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-variable (or discharge-variable) have to be assumed constant, e.g. storage zone area, dispersion coefficient, and exchange rate.

(2) Parameterization of transport clearly influenced model parameters related to other processes, 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. In addition, 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 would rather introduced additional uncertainties and that further insights provided by the model would be 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 purely "phenomenological". While the results of the cluster analysis are only valid in the investigated system, we see a potential for application of this method in other systems and added this aspect to the discussion:

> *"…, the cluster analysis helped to explore how the catchment and processes therein influenced concentration signals of mobilized contaminants. Particularly, the analysis helped to understand under what conditions and for which pollutant sharp-peaked chemographs, associated with high acute toxicity, can be expected. We therefore see a high potential of this type of analysis for the identification of influential factors for contaminant mobilization in other catchments, although these factors may not be universal but catchment-dependent."* (Lines 604-608)

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 sentence was revised and now reads:

> *"If degradation is incomplete, pesticides form transformation products (TPs) which in some cases may be equally or more mobile, persistent or toxic than their PCs (Hensen et al., 2020)."* (Lines 39-41)

L. 93: How often were grab samples taken?

Reply: We clarified the number of grab sample collections:

> *"Pesticide monitoring at G1 and G2 consisted of 5 manual sample collections during stationary flow conditions and 10 automated event samplings during discharge events."* (Lines 148-150)

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 added information to the description of the analytical methodology and provided a detailed assessment of analytical precision in the supplementary material:

“The following analytical methods were used for determining pesticide levels in the water samples. Analytical standards of boscalid (99.9%), penconazole (99.1%), metazachlor (99.6%), and flufenacet (99.5%) and the internal standards Diuron-D6 (99 %) and Terbutryn-D5 (98.5 %) already dissolved in acetonitrile (100 µg mL-1) were purchased from Sigma-Aldrich Chemie GmbH (Steinheim, Germany). Met-ESA (95 %) and met-OA (98.8 %) were received from Neochema (Bodenheim, Germany). Acetonitrile (LC-MS grade; VWR International GmbH, Darmstadt, Germany) was used as organic mobile phase in chromatography and for the preparation of stock solutions. Aqueous mobile phase was prepared with ultrapure water (Membra Pure, Germany; Q1:16.6 mΩ and Q2: 18.2 mΩ.

Preparation of environmental samples (approx. 1 liter) was done by filtering with a folded filter (type 113 P Cellulose ø 240 mm). Supernatant was spiked with the internal standard Diuron-D6 (10 µl of 10 mg L-1). Extraction procedure was a solid phase extraction (SPE). Cartridges (CHROMABOND® HR-X 6 mL/200 mg) were conditioned with 10 mL methanol and washed with 10 mL pure water. 90 µL of the extract were spiked with 10 µl of Terbutryn-D5 as an internal standard. Each sample was a double determination. Measurements of environmental samples were conducted with a Triple Quadrupole (Agilent Technologies, 1200 Infinity LC-System and 6430 Triple Quad, Waldbronn, Germany). Mobile phases were 0.01% formic acid (A) and acetonitrile (B) with a flow of 0.4 mL min^{-1}. Gradient was as follows: 0-1 min (10% B), 1-11 min (10-50% B), 11-18 min (50-85% B), 18-21 min (85-90% B), 21-24 min (90% B), 24-26 min (90-10% B) and 26-30 (10% B). A NUCLEODUR® RP-C18 (125/2; 100-3 µm C18 ec) column (Macherey Nagel, Düren, Germany) was used as stationary phase with a set oven temperature of T = 30°C. Calibration curve were prepared in pure water. The linearity was evaluated by preparing three curves with ten calibration points in the range 1 - 500 µg/L. The standard curves were then extracted according to the protocol and analyzed using LC-MS/MS. The calculated linear regression values (R^2) were very good with R^2-values > 0.999. The linearity between peak area and concentration of substances were obtained in a range of 0 - 5 µg L^{-1}. Hence limits of detection (LOD) and quantitation (LOQ) were calculated with DINTEST (2003) according to DIN 32645 considering an enrichment factor of 5000. LOD and LOQ amounted to 0.4 and 1.3 ng L^{-1} (boscalid), 0.3 and 0.9 ng L^{-1} (penconazole), 0.3 and 1.2 ng L^{-1} (metazachlor), 0.4 and 1.3 ng L^{-1} (flufenacet) as well as 0.6 and 2.2 ng L^{-1} (met-ESA) and 0.5 and 1.6 ng L^{-1} (met-OA) considering an enrichment factor of 5000. A detailed analysis of measurement precision can be found in the supplementary material.” (Lines 178-207)

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

Reply: The high number of digits resulted from the calculation of LODs and LOQs. We rounded these values so that they corresponded to the analytical precision (see above).

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 was 3.6.1 and the algorithm used is specified:

> *"The analysis was done using the software R (R Core Team, 2019) (version 3.6.1) using the 'pam' (partitioning around medoids) function from the 'cluster'-package (version 2.1.0) (Maechler et al., 2019)."* (Lines 219-220)

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: The linear regression model was removed. See response to referee #2.

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

Reply: The equation is given in the revised manuscript:

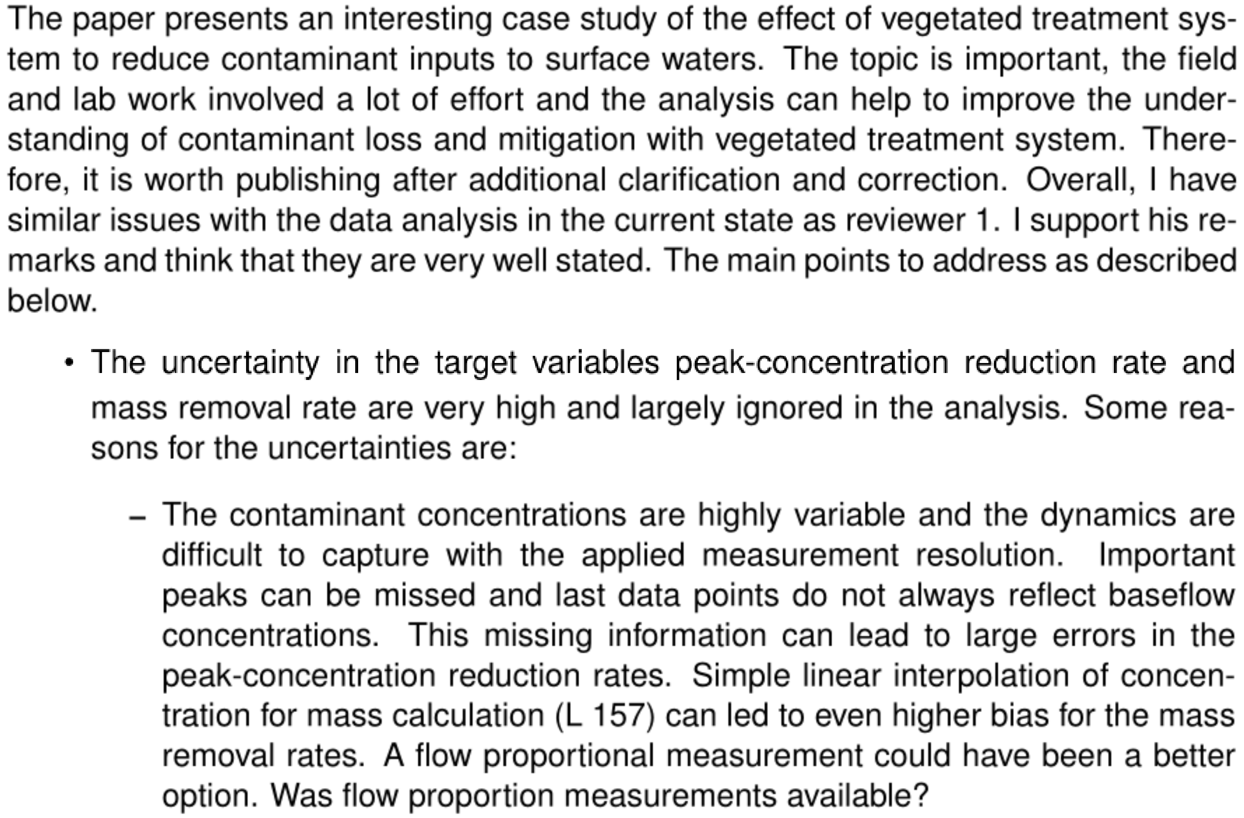
"$$W_B = \frac{Q_{in,mean} - Q_{out,mean}}{Q_{in,mean}} \cdot 100\ \%\ , \qquad (4)$$

where $Q_{in,mean}$ *and* $Q_{out,mean}$ *are the discharge at G1 and G2, respectively, averaged over the duration of the sampling procedure at both gauges.* W_B *was positive, if more water entered the wetland than left the wetland during the sampling procedure, and negative in the opposite case."* (Lines 259-262)

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: As sated above, we excluded the regression model and hence also the exclusion of outlyers.

Response to comments by Referee #2



The paper presents an interesting case study of the effect of vegetated treatment system to reduce contaminant inputs to surface waters. The topic is important, the field and lab work involved a lot of effort and the analysis can help to improve the understanding of contaminant loss and mitigation with vegetated treatment system. Therefore, it is worth publishing after additional clarification and correction. Overall, I have similar issues with the data analysis in the current state as reviewer 1. I support his remarks and think that they are very well stated. The main points to address as described below.

- The uncertainty in the target variables peak-concentration reduction rate and mass removal rate are very high and largely ignored in the analysis. Some reasons for the uncertainties are:
    - The contaminant concentrations are highly variable and the dynamics are difficult to capture with the applied measurement resolution. Important peaks can be missed and last data points do not always reflect baseflow concentrations. This missing information can lead to large errors in the peak-concentration reduction rates. Simple linear interpolation of concentration for mass calculation (L 157) can led to even higher bias for the mass removal rates. A flow proportional measurement could have been a better option. Was flow proportion measurements available?

Reply: We adopted our measurement setup to inculde both aspects of contaminant mitigation, i.e. peak concentration decrease and mass removal and we performed a cluser analysis to group chemograph shapes. This required measurements of original contaminant concentrations. We agree that for mass removal flow proportion sampling schemes are more accurate. We included a section in the discussion dealing with the measurement set up and associated uncertainties:

> *"Regarding chemographs and calculation of R_C, uncertainties arose from timing and frequency of sampling and analytical error, and additionally from discharge measurement when calculating masses and R_M. Analytical methods used in this study usually produced very consistent results so that variability in concentrations of parent compounds in duplicate samples was low (sd < 10 %). However, in individual samples collected at G1 analytical variability was elevated for met-ESA and met-OA (Figure 3), reducing confidence in concentrations and the derived measures R_C and R_M of TPs in the affected chemographs (E2, E5, E7, E8, E9). Uncertainty related to timing and frequency of sampling can hardly be quantified but certainly depends on how well the sampling intervals captured variability in concentrations during flood events and how well the time lag between upstream and downstream sampling matched the residence time of solutes in the wetland. Lefrancq et al. (2017) assessed the effect of sampling frequency in pesticide monitoring data*

> *collected during runoff from a single vineyard and found that acute toxicity of pesticide flushes was underestimated up to 4-times when calculated from event means and up to 30-times when calculated from random samples. Although these data were collected on the plot scale and we assume that variability in our catchment is lower due to longer flow paths and mixing processes on the catchment scale, uncertainty of the chemographs in our study could have been reduced by increasing sampling frequency. Regarding the timing of upstream and downstream sampling, there is evidence that water residence time in the wetland was in fact shorter than one hour. The observation that for quickly responding compounds, such as boscalid, concentration in the first sample at G2 was often elevated compared to the first sample at G1 indicates that the contaminant flush hat already reached G2 when sampling started. This did not influence determination of* $C_{out,max}$ *and* R_C *in the outlet of the wetland, as concentrations were still rising from the first to the second sample (Figure 3). However, effects on* M_{out} *were higher, since a relevant fraction of contaminant mass leaving the wetland was not registered and thereby caused overestimation of* R_M *(Figure S1). Another source of uncertainty exclusively affecting contaminant mass and not concentrations was the use of different gauging systems at G1 and G2. Different shapes of the measurement cross-section (triangular at G1 and rectangular at G2) caused G2 to be less precise and water imbalances on the event scale, particularly when flow was low. Summarizing the setup constraints above, we have high confidence that the experimental setup produced realistic chemograph shapes and captured peak concentration reasonably well, but are less confident regarding contaminant loads."* (Lines 524-547)

- The timing of the last application of the investigated substances is completely neglected. However, more pronounced peaks are to be expected shortly after the application. Standardizing the concentrations or considering relative reduction rate does not completely solve this issue.

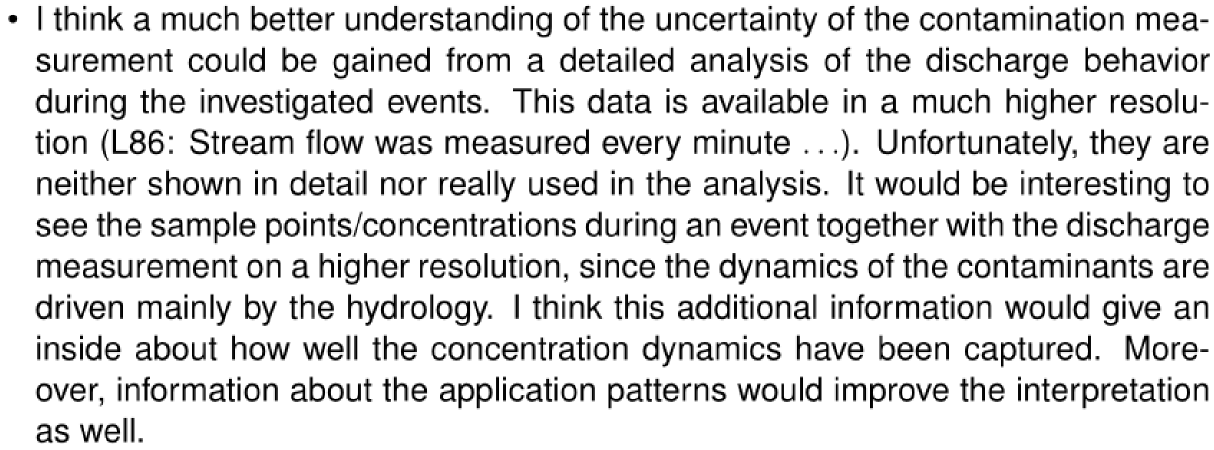
Reply: We agree that detailed information on application rates would be beneficial for interpretation of the pesticide signals emerging from the catchment. We do not have data on timing and applied amounts, but included this aspect in the discussion:

> *"Cluster C was composed of chemographs of all compounds but mainly from events E2 and E4 indicating event dependence. Two aspects were found to support this idea. First, there was a secondary discharge peak in event 2 that did not contribute much in terms of contaminant concentration but rather caused dilution and produced particularly flat chemograph tails. Second, peaks of herbicides and TPs were less delayed compared to fungicides. This may be the result of recent herbicide application and active surface runoff in the flat valleys. Timing of pesticide application was identified as the main driver of pesticide export by Imfeld et al. (2020) who performed a cluster analysis on rainfall data from a headwater vineyard catchment. Based on the magnitude of discharge and amount of mobilized contaminants (concentration of metazachlor ≈*

> *10 µg L^{-1}), both explanations seem plausible in event E2. Event E4, however, did not show particularly high herbicide concentration nor a secondary discharge peak. Although it is obvious that chemograph shapes in cluster C differed from the other clusters, unfortunately, the responsible factors remain unclear."* (Lines 584-593)

And also in lines 599-603:

> *"The unclear interpretation of cluster C suggest that we missed important factors for the formation of chemographs. In fact, variables like spatial distribution of rainfall or pesticide application rates and timing (Imfeld et al., 2020) and possibly other factors likely influenced chemograph shapes. Knowing all these variables would not change the results produced by the clustering algorithm but rather increase our ability to interpret them. Nevertheless, the cluster analysis helped to explore how the catchment and processes therein influenced concentration signals of mobilized contaminants."* (Lines 601-608)



- I think a much better understanding of the uncertainty of the contamination measurement could be gained from a detailed analysis of the discharge behavior during the investigated events. This data is available in a much higher resolution (L86: Stream flow was measured every minute . . .). Unfortunately, they are neither shown in detail nor really used in the analysis. It would be interesting to see the sample points/concentrations during an event together with the discharge measurement on a higher resolution, since the dynamics of the contaminants are driven mainly by the hydrology. I think this additional information would give an inside about how well the concentration dynamics have been captured. Moreover, information about the application patterns would improve the interpretation as well.

Reply: We added a Figure (Figure 3) showing discharge dynamics and pesticide responses in detail (see above) and included discharge dynamics in the discussion:

> *"Cluster D included chemographs of both herbicides and TPs and presented a clear peak that was often defined by a single sample 2 h after the beginning of the event. In contrast to cluster A, cluster D was characterized by a single sharp discharge peak (except in event E7 where a second peak occurred shortly after the first) and mainly included chemographs during periods of low flow. Our interpretation is that cluster D represented flow events in which no dilution of herbicide and TP fluxes by fungicide fluxes or secondary discharge peaks occurred. Low pre-event discharge in cluster D compared to cluster A may indicate low water levels which may have caused a slower response as no enriched pre-event water was released from the soils in the valleys. "* (Lines 594-600)

- The regression analysis is done rather poorly and the procedure is neither well explained nor well presented.
  - It is not shown that the condition for a simple multiple linear regression are fulfilled, since the results are not validated or at least this information is not shown. I would at least expect a classical residual analysis in the supporting information.

Reply: We are thankful for this concern. After careful consideration of the statistical constraints of our data set, particularly with respect to the following points, we decided to replace the regression analysis by a qualitative comparison of mitigation in the wetland to compound properties, discharge dynamics and chemograph shape parameters. The results are shown in Figure 5 and Figure 6 and discussed as follows:

"4.3 Mitigation efficiency and chemograph shape

4.3.1 Peak concentration reduction

We hypothesized that peak concentration reduction in the VTS will be highest for chemographs with the sharpest peaks, i.e. for the chemographs that were most sensitive to dispersion. And indeed we found a systematic relationship between R_C *and both* i_{DS} *and chemograph clusters. Although the relationship of clusters and* R_C *largely reflected the relationship between* R_C *and* i_{DS}*, it is surprising that* R_C *was clearly highest in cluster D and not in cluster C which presented better defined peaks and slightly higher* i_{DS} *per cluster (Figure 4f). Critical inspection of input chemographs shows that in several chemographs of TPs (met-ESA and met-OA in event E4 and met-OA in event E8) elevated concentrations in the last samples exhibited high analytical errors and did not appear in the outlet chemograph. These dubious samples caused low* i_{DS} *but substantial* R_C *and thus contributed to variability in* i_{DS} *despite high values of* R_C *in cluster D. We therefore do not consider the deviation from the expected cluster ordering contradictory but to result from increased uncertainty in cluster D as mentioned earlier. In contrast, the hypothesized relationship between RC and chemograph shape was demonstrated for both* i_{DS} *and chemograph clusters, the latter of which also integrates shape aspects that go beyond* i_{DS}*, e.g. timing of peaks. Overall, the values of* R_C *found in our study compare with field data from vegetated buffers (Bundschuh et al., 2016; Stehle et al., 2011) and are in the range of those found in vegetated stream mesocosms by Elsaesser et al. (2011) and Stang et al. (2014) who both attributed most of the observed peak reduction to dispersion.*

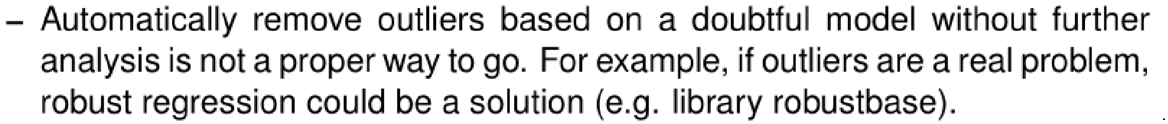
In addition, we found relationships between and R_C *and discharge dynamics, i.e.* Q_{mea} *and ratio of* Q_{max} *to* Q_{mean}*. The influence of discharge on* R_C *may be two-fold. First, increasing flow reduced residence time and hydraulic efficiency, i.e. short circuiting reduced the potential for dispersion and interaction with wetland sediments or plants. Second, the fact that chemographs of events* *with high Q_{max} to Q_{mean} ratios were attributed to cluster D suggests that discharge dynamics influenced the shape of the chemograph at the wetland inlet. This means, the influence of discharge may also be indirect by promoting the formation of sharp-peaked chemographs with high potential for peak reduction.*

In contrast to other studies, we did not find clear relationships of R_C to and physiochemical properties of compounds such as sorption affinity (Stehle et al., 2011; Vymazal and Březinová, 2015) or solubility (Bundschuh et al., 2016). The absence of such relationships may partially be due to the low number of different target compounds in our study (n=6). However, given the short time lag between sampling at the inlet and outlet of the wetland ($\Delta t = 1h$), it seems logical that no relevant sorption or degradation occurred within this period. For comparison, in batch experiments by Gaullier et al. (2018) adsorption equilibrium for boscalid (compound with second highest Kf_{OC} in our study) was only reached after 24 h. Despite the relatively narrowly confined R_C values of the parent compounds, we do not consider physiochemical compound properties as major drivers of R_C in our VTS.

4.3.2 Contaminant mass removal

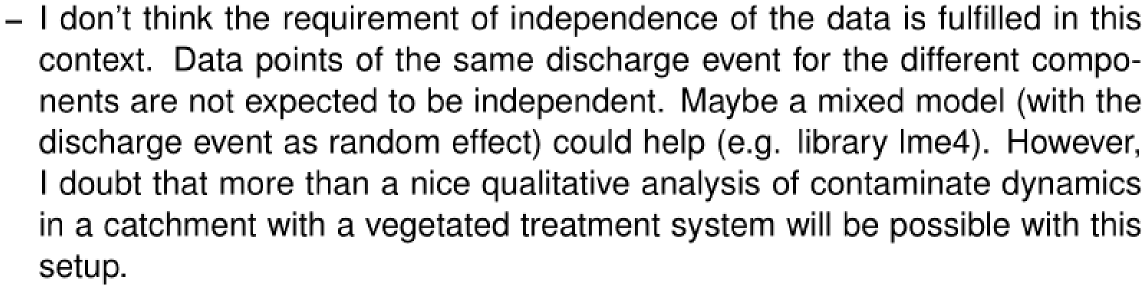
For R_M we found a different pattern among the chemograph clusters than for R_C. R_M was apparently higher in clusters A and D than in clusters B and C. However, the clusters indicating substantial mass removal were those with increased uncertainty regarding compound mass. Cluster A often showed relevant mass flux at the end of sampling (and presumably beyond) which we did not account for. Cluster D contained dubious data points of TPs and poorly defined peaks outside the periods of high sampling frequency. In addition, due to overestimation of solute travel time in the wetland in the monitoring setup, the rising limp of the mass flux signal at G2 was often not adequately captured by the sampling scheme, causing underestimation of downstream event mass and overestimation of mass loss. In absence of any clear relationship with compound properties, discharge dynamics or chemograph shape, this suggests that the assessment of contaminant masses was subject to systematical errors and that the apparent mass loss found in our study should therefore not be over-interpreted.

In earlier studies, Lange et al. (2011) and Schuetz et al. (2012) observed a 15-30 % mass loss of the fluorescent tracer sulforhodamine-B in the wetland subsection of the current VTS. These results indicate a general potential for sorption of organic compounds in this system, but represent an earlier succession state of the wetland and stationary flow conditions with much longer residence times. Also in the current VTS kinetic sorption of contaminants may have occurred but sorption equilibrium was certainly not reached (Gaullier et al., 2018). Thus the effect of sorption did not reach its full potential. In fact, other studies reported limited mass removal in wetlands with comparable residence times. Ramos et al. (2019) did not find relevant R_M in two surface flow *wetlands with residence times between 45 min and 6 h in England. In contrast, Passeport et al. (2013) found R_M between 45 % and 96 % in a constructed wetland with a residence time of 66.5 h. However, their contaminant mass loss coincided with loss of water (45 %). Mesocosm experiments by Elsaesser et al. (2011) and Stang et al. (2014) showed strong concentration reduction but only very limited and temporary mass removal at residence times of a few hours. In summary, these findings suggest that the potential for mass removal in wetland systems like the one studied here is rather limited. However, wetlands have been shown to reduce contaminant mass, when residence times are sufficiently long (Gregoire et al., 2009) or when operated in batch mode (Tournebize et al., 2017; Moore et al., 2000; Maillard et al., 2016).”* (Lines 609-661)



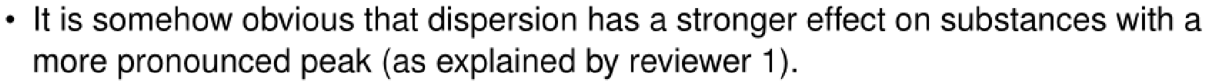
- Automatically remove outliers based on a doubtful model without further analysis is not a proper way to go. For example, if outliers are a real problem, robust regression could be a solution (e.g. library robustbase).

Reply: As stated above, we entirely refrained from removing outliers in the revised manuscript.



- I don’t think the requirement of independence of the data is fulfilled in this context. Data points of the same discharge event for the different components are not expected to be independent. Maybe a mixed model (with the discharge event as random effect) could help (e.g. library lme4). However, I doubt that more than a nice qualitative analysis of contaminate dynamics in a catchment with a vegetated treatment system will be possible with this setup.

Reply: See also above. We are grateful for this comment and the helpful suggestion to perform a qualitative analysis instead of a questionable regression model.



- It is somehow obvious that dispersion has a stronger effect on substances with a more pronounced peak (as explained by reviewer 1).

Reply: We agree with referee #2 that it is obvious that dispersion affects sharp peaks stronger than flattened signals. However, as already discussed in our response to reviewer #1, it is not adequately stressed in existing literature on wetland contaminant mitigation.

- Although the clustering is done correctly, the connection with the discharge events is not well elaborated. Moreover, they are other clustering algorithms, which might be more robust (e.g. k-medoids. hierarchical Clustering). In L 207 it is written:"With the exception of cluster B which rather represented similar events (event 1 and event 4 in Fig. 2), overall clustering was controlled by similar behavior of contaminant groups." What was special by the event 1 and 4? Are these really exceptions? The contaminant groups seem to be important, however, I think the discharge dynamic and the application timing are important as well. Maybe it would also been interesting to cluster the discharge events. This data are also available in a higher resolution.

Reply: We changed the clustering algorithm to k-medoids which should increase robustness against outliers:

> *"Identification of patterns in input chemographs was done by k-medoids cluster analysis - a variation of the commonly applied k-means algorithm. Both approaches partition the elements of a dataset into a predefined number k of clusters by attributing the elements to the cluster with the nearest cluster center. Optimal clustering is achieved by iteratively updating cluster centers and minimizing distance between data points and cluster centers. K-medoids differs from k-means as it uses existing points (medoids) as cluster centers instead of means and is considered more robust against extreme values and outliers (Han et al., 2012). A total of 58 concentration sequences was included in the analysis, consisting of 10 sequences per target compound, except for flufenacet which did not exceed LOQ in two events. Prior to cluster analysis, data was normalized by the maximum of each chemograph to promote that clustering represented shape, rather than differences in absolute concentration. The analysis was done using the software R (R Core Team, 2019) (version 3.6.1) using the 'pam' (partitioning around medoids) function from the 'cluster'-package (version 2.1.0) (Maechler et al., 2019). We tested clustering for k ranging between 2 and 10, the final number was determined by both visual inspection of the clusters and assessment of explanatory benefit per additional cluster (elbow method). As a result we found that k=4 resulted in the best partition."* (Lines 211-222)

Change of clustering algorithm caused minor differences in attribution of chemographs to clusters as described by Figure 4 and the following paragraph:

> *"Cluster A (Figure 4) was characterized by absence of a clear peak during the first two hours of sampling but elevated concentrations during later times, resulting in low i_{DS}. Cluster B showed a quick response, i.e. concentrations increased sharply within the first 30 minutes. Concentrations were the highest of all clusters and still elevated in the last sample compared to pre-event levels. Cluster C was characterized by a clear peak within the first two hours and a low tailing and was*

> *the cluster with highest median i_{DS}. Cluster D showed the most inconsistent pattern and maximum concentrations appeared later compared to clusters B and C. A relatively clear pattern was evident in the attribution of compounds to the clusters. Chemographs of the fungicides boscalid and penconalzole were mainly assigned to cluster B, while the herbicides and the TPs were assigned to the remaining three clusters. Cluster A was composed of herbicide and TP chemographs, particularly from events with multiple discharge peaks. Cluster D represented chemographs of herbicides and TPs mainly during the events E5 to E8 which were all characterized by sharp discharge peaks during periods of generally low flow (Figure 3). Almost all chemographs of the events E2 and E4 were attributed to cluster C."* (Lines 343-353)

We put more emphasis on the discussion of chemograph clusters, including the role of discharge conditions, and particularities of certain events:

> *"Despite highly variable application patterns, our cluster analysis resulted in four groups with similar chemograph shape. Many factors have been shown in literature to influence the mobilization of pesticides in catchments, including catchment properties, event properties and physiochemical compound properties. As catchment properties we here consider factors associated with runoff generation such as catchment geometry, terrain slopes, and in particular the delineation of areas where different compounds were applied. The interplay of these factors defines hydrological activity and connectivity (i.e. by shortcuts like roads and drainage pipes) of critical source areas for different compounds (Doppler et al., 2012; Gomides Freitas et al., 2008). Event properties include intensity and dynamics of rainfall (Imfeld et al., 2020) and subsequent runoff (Doppler et al., 2014). Relevant physiochemical compound properties are e.g. mobility and degradability (Gassmann et al., 2015).*
>
> *These properties are reflected to varying degrees in the results of the cluster analysis. Cluster A was characterized by a quick response and a concentration plateau towards the end of sampling and was mainly composed of TPs. The fact that concentration maxima in cluster A were retarded compared to fungicides (cluster B), although their parent compounds were applied closer to the stream in the flat valley bottoms, suggests that they were transported with a slower flow components. Due to flatter terrain, surface runoff played a less important role and the main transport pathway was subsurface flow. Where fields were undrained, however, transit time of water from the infiltration point to the stream would likely exceed the temporal scale of event sampling. Most of the water reaching the stream from the fields in the valley during discharge events would therefore be pre-event water, enriched in TPs formed in the soil, corresponding to the formation site of TPs of the chloracetamide herbicides to which metazachlor belongs (Mersie et al., 2004). Seepage of pre-event TP-rich water thus explains the immediate response of chemographs in cluster A. The quick response was often followed by a local concentration minimum between samples 2 and 5, i.e. between 30 min and 6 h after sampling was initialized.*

Coincidence of this minimum with concentration peaks of fungicides might suggest dilution of TP concentration by mixing with event water carrying high loads of fungicides but less TPs of metazachlor.

Cluster B represented differences between fungicides and the remaining compounds. Considering land use distribution in the studied catchment, it is unclear whether this partition reflects different compound properties or catchment properties or both. The fact that concentration in cluster B quickly increased with discharge (within 30 minutes) is in line with fast transport from the vineyard terraces to the stream via roads and drainage pipes as described by Gassmann et al. (2012) for suspended solids in the studied catchment. Along such preferential pathways, compound properties, such as sorption affinity, may be less important (Gomides Freitas et al., 2008) compared to e.g. percolation through the soil with intense contact to sorption sites in the soil matrix. Moreover, fungicides are applied by sprayers into the foliage and can drift to e.g. paved surfaces from which they can be quickly mobilized by subsequent rainfall (Lefrancq et al., 2013). We therefore hypothesize that cluster B was mainly produced by surface flushing and fast transport pathways of fungicides. This explained the quick rise and subsequent decline in concentrations (concurrent with plateaus produced by slower flow components in cluster A).

Cluster C was composed of chemographs of all compounds but mainly from events E2 and E4 indicating event dependence. Two aspects were found to support this idea. First, there was a secondary discharge peak in event 2 that did not contribute much in terms of contaminant concentration but rather caused dilution and produced particularly flat chemograph tails. Second, peaks of herbicides and TPs were less delayed compared to fungicides. This may be the result of recent herbicide application and active surface runoff in the flat valleys. Timing of pesticide application was identified as the main driver of pesticide export by Imfeld et al. (2020) who performed a cluster analysis on rainfall data from a headwater vineyard catchment. Based on the magnitude of discharge and amount of mobilized contaminants (concentration of metazachlor ≈ 10 µg L^{-1}*), both explanations seem plausible in event E2. Event E4, however, did not show particularly high herbicide concentration nor a secondary discharge peak. Although it is obvious that chemograph shapes in cluster C differed from the other clusters, unfortunately, the responsible factors remain unclear.*

Cluster D included chemographs of both herbicides and TPs and presented a clear peak that was often defined by a single sample 2 h after the beginning of the event. In contrast to cluster A, cluster D was characterized by a single sharp discharge peak (except in event E7 where a second peak occurred shortly after the first) and mainly included chemographs during periods of low flow. Our interpretation is that cluster D represented flow events in which no dilution of herbicide and TP fluxes by fungicide fluxes or secondary discharge peaks occurred. Low pre-event discharge in cluster D compared to cluster A may indicate low water levels which may have

> *caused a slower response as no enriched pre-event water was released from the soils in the valleys.*
> *The uncertain interpretation of cluster C suggest that we missed important factors for the formation of chemographs. In fact, variables like spatial distribution of rainfall or pesticide application rates and timing (Imfeld et al., 2020) and possibly other factors likely influenced chemograph shapes. Knowing all these variables would not change the results produced by the clustering algorithm but rather increase our ability to interpret them. Nevertheless, the cluster analysis helped to explore how the catchment and processes therein influenced concentration signals of mobilized contaminants."* (Lines 522-608)

Detailed comments

L100: Fig. 2: I guess the discharge shown in Fig. 2 is from G1. This should be included in the description.
Reply: We replaced Figure 2 by a more detailed version (Figure 3) with a proper legend (see above).

L 108: Overall, herbicides have been also shown to be very persistent. For examples, atrazine has been detected after 10 years without application.
Reply: This was not meant to be a general statement. These lines (also s. next comment) just sum up, how the selected compounds are classified according to the Pesticide Properties Data Base (Lewis et al., 2016). We clarified this in the revised manuscript:

> *"In this study, we focused on 6 target compounds including the fungicides boscalid and penconazole, the herbicides metazachlor and flufenacet, and the TPs metazachlor sulfonic acid (met-ESA) and metazachlor oxalic acid (met-OA). Selected physicochemical properties of the target compounds are listed in Table 1. According to the Pesticide Properties Data Base (Lewis et al., 2016) the contaminants can be classified as low (boscalid) to moderately soluble in water, very mobile (TPs) to slightly mobile (fungicides). The target fungicides are considered moderately fast degradable in the water phase and persistent in soils, while the target herbicides are considered stable in the water phase and non-persistent in soils. TPs of metazachlor are considerably more persistent in soil than their PC. The fungicides are considered stable with respect to hydrolysis but degradable via photolysis, while the herbicides are stable regarding both."* (Lines 130-137)

L 110: Azole pesticides are also persistent as indicated by many studies (e.g. https://doi.org/10.1016/j.envint.2020.105708)
Reply: (s. above)

L 131: What is the accuracy and precision of the method? Has the analytical method validated?
Reply: We provided additional information about the analytical procedure in the method section and added a detailed analysis of the measurement precision in the supplementary material.

> *“The following analytical methods were used for determining pesticide levels in the water samples. Analytical standards of boscalid (99.9%), penconazole (99.1%), metazachlor (99.6%), and flufenacet (99.5%) and the internal standards Diuron-D6 (99 %) and Terbutryn-D5 (98.5 %) already dissolved in acetonitrile (100 µg mL-1) were purchased from Sigma-Aldrich Chemie GmbH (Steinheim, Germany). Met-ESA (95 %) and met-OA (98.8 %) were received from Neochema (Bodenheim, Germany). Acetonitrile (LC-MS grade; VWR International GmbH, Darmstadt, Germany) was used as organic mobile phase in chromatography and for the preparation of stock solutions. Aqueous mobile phase was prepared with ultrapure water (Membra Pure, Germany; Q1:16.6 mΩ and Q2: 18.2 mΩ.*
>
> *Preparation of environmental samples (approx. 1 liter) was done by filtering with a folded filter (type 113 P Cellulose ø 240 mm). Supernatant was spiked with the internal standard Diuron-D6 (10 µl of 10 mg L-1). Extraction procedure was a solid phase extraction (SPE). Cartridges (CHROMABOND® HR-X 6 mL/200 mg) were conditioned with 10 mL methanol and washed with 10 mL pure water. 90 µL of the extract were spiked with 10 µl of Terbutryn-D5 as an internal standard. Each sample was a double determination. Measurements of environmental samples were conducted with a Triple Quadrupole (Agilent Technologies, 1200 Infinity LC-System and 6430 Triple Quad, Waldbronn, Germany). Mobile phases were 0.01% formic acid (A) and acetonitrile (B) with a flow of 0.4 mL min^{-1}. Gradient was as follows: 0-1 min (10% B), 1-11 min (10-50% B), 11-18 min (50-85% B), 18-21 min (85-90% B), 21-24 min (90% B), 24-26 min (90-10% B) and 26-30 (10% B). A NUCLEODUR® RP-C18 (125/2; 100-3 µm C18 ec) column (Macherey Nagel, Düren, Germany) was used as stationary phase with a set oven temperature of T = 30°C. Calibration curve were prepared in pure water. The linearity was evaluated by preparing three curves with ten calibration points in the range 1 - 500 µg/L. The standard curves were then extracted according to the protocol and analyzed using LC-MS/MS. The calculated linear regression values (R^2) were very good with R^2-values > 0.999. The linearity between peak area and concentration of substances were obtained in a range of 0 - 5 µg L^{-1}. Hence limits of detection (LOD) and quantitation (LOQ) were calculated with DINTEST (2003) according to DIN 32645 considering an enrichment factor of 5000. LOD and LOQ amounted to 0.4 and 1.3 ng L^{-1} (boscalid), 0.3 and 0.9 ng L^{-1} (penconazole), 0.3 and 1.2 ng L^{-1} (metazachlor), 0.4 and 1.3 ng L^{-1} (flufenacet) as well as 0.6 and 2.2 ng L^{-1} (met-ESA) and 0.5 and 1.6 ng L^{-1} (met-OA) considering an enrichment factor of 5000. A detailed analysis of measurement precision can be found in the supplementary material.”* (Lines 178-207)

L 145: Why is the cluster analysis important for the calculation of the dispersion sensitivity index? The index could also be calculated without clustering.

Reply: We acknowledge that this phrasing was imprecise. We made the connection clear in the revised manuscript.

L 191: From which mean? Do you mean 2 standard deviation from the prediction?
Reply: No outliers were removed in the revised manuscript.

L 205: It doesn't make sense to talk about a peak in Cluster C ("T peak = 6h"). Not even the mean has a peak there.
Reply: This is true and was avoided in the revised manuscript.

L 212: The surface runoff from the elevated vineyard has also to flow through the lower terrain slope to reach the river, expect that there are other shortcuts (streets, drains). See also reviewer 1).
Reply: We thank referee #2 for this comment and put stronger emphasis on the role of artificial drainage and other shortcust inside the catchment in the revised manuscript both in the description of the study site:

> *"Large parts of the catchment are drained by a sub-surface pipe network (Figure 1) connecting vineyards and paved roads to the main channel in the valley. This drainage network causes fast downstream transport of storm water and suspended sediments (Gassmann et al., 2012). In addition, fields in the valley bottoms are drained by a secondary network of smaller and usually shorter field drains that are either connected to the primary drainage network or directly connected to the stream (Schuetz et al., 2016)."* (Lines 106-111)

We also included this aspect in the discussion of contaminant mobilization, particularly in the case of cluster B:

> *"Cluster B represented differences between fungicides and the remaining compounds. Considering land use distribution in the studied catchment, it is unclear whether this partition reflects different compound properties or catchment properties or both. The fact that concentration in cluster B quickly increased with discharge (within 30 minutes) is in line with fast transport from the vineyard terraces to the stream via roads and drainage pipes as described by Gassmann et al. (2012) for suspended solids in the studied catchment. Along such preferential pathways, compound properties, such as sorption affinity, may be less important (Gomides Freitas et al., 2008) compared to e.g. percolation through the soil with intense contact to sorption sites in the soil matrix. Moreover, fungicides are applied by sprayers into the foliage and can drift to e.g. paved surfaces from which they can be quickly mobilized by subsequent rainfall (Lefrancq et al., 2013). We therefore hypothesize that cluster B was mainly produced by surface flushing and fast transport pathways of fungicides. This explained the quick rise and subsequent decline in concentrations (concurrent with plateaus produced by slower flow components in cluster A)."* (Lines 574-583)

L 315: I do not understand the explanatory power for the different variable. Are they calculated by a univariate analyses? At least for me, the R-Output would be much easier to interpret.
Reply: The regression analysis is no longer part of the manuscript.

Pesticide peak concentration reduction in a small vegetated treatment system controlled by chemograph shape

Jan Greiwe[1], Oliver Olsson[2], Klaus Kümmerer[2], Jens Lange[1]

[1]Hydrology, University of Freiburg, Fahnenbergplatz, 79098 Freiburg, Germany
[2]Institute of Sustainable and Environmental Chemistry, Leuphana University of Lueneburg, Scharnhorststr. 1, 21335 Lueneburg, Germany

Correspondence to: Jan Greiwe (jan.greiwe@hydrology.uni-freiburg.de)

Abstract. Pesticides may impact aquatic ecosystems when entering water bodies. Measures for mitigation against pesticide inputs include vegetated treatment systems (VTS). Some of these systems have very short hydraulic retention time (< 1 h) but nevertheless manage to effectively reduce peak concentrations of contaminants as a result of dispersion. We hypothesize that the effect of dispersion on contaminant mitigation in VTS depends on the shape of the contaminant input signal chemograph which in turn is related to factors affecting contaminant mobilization in the contributing catchment. In order to test this hypothesis we grouped chemographs of 6 contaminants ~~mobilized~~originating from a viticultural catchment during 10 discharge events into clusters according to chemograph shape. We ~~T~~then compared peak concentration reduction and mass removal in a downstream VTS both among clusters and in terms of compound properties and discharge dynamics. We found that chemograph clusters reflected combined effects of contaminant source areas, transport pathways, and discharge dynamics. While mass loss was subject to major uncertainties, peak concentration reduction rate was clearly related to chemograph clusters and dispersion sensitivity. ~~. We hypothesize that this is not solely the result of contaminant and VTS properties but also related to the shape of the contaminants input chemograph, i.e. its sensitivity to dispersion. In order to test this hypothesis we performed a cluster analysis on the chemographs of contaminants mobilized in response to rainfall events in a viticultural catchment and derived a measure for dispersion sensitivity which we included into multiple linear regression analysis. The resulting measure was then incorporated into multiple linear regression models for description of contaminant mitigation in a VTS located at the catchment outlet. We found that the mobilization clusters reflected the source areas of the contaminants and that dispersion sensitivity was the dominant explanatory variable for the reduction of contaminant peak concentration.~~ These findings suggest that mitigation of acute toxicity in VTS is stronger for compounds with sharp-peaked chemographs whose formation is related to the contributing catchment and can be analyzed by chemograph clustering.

1 Introduction

Use of pesticides is beneficial for agricultural productivity. However, w~~W~~hen pesticides reach surface water bodies, they threaten aquatic organisms (Zubrod et al., 2019). Effects of pesticides on aquatic ecosystems include a reduction of species richness of invertebrates (Beketov et al., 2013) as well as microorganisms (Fernández et al., 2015). Unintended export of pesticides from the application site to water bodies can happen in particulate form via erosion (Oliver et al., 2012; Taghavi et al., 2011) or in dissolved form via surface runoff, drainage pipes, spray drift or leakage to groundwater (Reichenberger et al., 2007) and subsequent exfiltration.

In the environment pesticides are subject to degradation by both abiotic (e.g. hydrolysis, photolysis) and biotic (e.g. plant metabolism, microbial degradation) processes (Fenner et al., 2013). If degradation is incomplete, pesticides form transformation products (TPs) which in some cases may be equally or more mobile, persistent or toxic than their PCs (Hensen et al., 2020). Strongest mobilization of pesticides is usually associated with the first significant rainfall after application and fast discharge components (Doppler et al., 2012; Olsson et al., 2013; Lefrancq et al., 2017), e.g. runoff from non-target areas like roads (Lefrancq et al., 2013). Mobilization dynamics of TPs usually differ from ~~that~~ those of their parent compounds (PCs) in terms of ~~both~~ source areas and export pathways. ~~as t~~The formation of TPs may happen on larger time scales and TPs ~~mostly~~ usually have different physicochemical properties than their PCs (Gassmann et al., 2013). The specific transformation and further degradation of a contaminant largely depends on the interplay of the contaminant's mobility and degradability as well as site characteristics (Gassmann et al., 2015). Both mobility and degradability can vary over multiple orders of magnitude for different contaminants. However, water and soil half-lives are at least in the order of several days or weeks for most pesticides (Lewis et al., 2016).

Pesticide pollution can be mitigated by vegetated treatment systems (VTS) located between source areas and receiving water bodies (Vymazal and Březinová, 2015; Stehle et al., 2011; Gregoire et al., 2009). Such systems temporally retain polluted waters and thus provide space, time and favorable conditions for degradation processes~~ to happen~~. VTSs studied in literature include very different types of systems (Lange et al., 2011), including vegetated ditches or detention ponds with hydraulic ~~retention~~ residence times (HRT) ranging in the order of minutes to several hours (Bundschuh et al., 2016; Elsaesser et al., 2011; Ramos et al., 2019) or constructed wetlands in which HRT may reach several weeks (Maillard and Imfeld, 2014), particularly when operated in batch mode (Tournebize et al., 2017; Maillard et al., 2016; Moore et al., 2000). The term pesticide mitigation can refer to contaminant mass removal (R_M) or peak concentration reduction (R_C). While mass removal ~~While contaminant mass loss (R_M)~~ is mainly observed in systems with longer HRT and affects ~~long-term~~permanent toxicity, peak concentration reduction also happens in systems with short HRT ~~have been shown to efficiently reduce peak concentration (R_C) and the associated~~ where it reduces acute toxicity ~~associated with compoundscontaminants~~ (Bundschuh et al., 2016; Elsaesser et al., 2011; Stehle et al., 2011).

During longitudinal transport in streams or wetlands, peak concentration reduction does not necessarily involve degradation but can solely be the result of enhanced dispersion due to the presence of obstacles such as plants (Elsaesser et al., 2011) and temporary removal from the water phase by reversible sorption. Mitigation properties therefore constantly change due to wetland succession (Schuetz et al., 2012). Regardless of whether VTSs target concentration reduction or mass removal, mitigation efficiency is usually associated with physicochemical properties of target compounds (Vymazal and Březinová, 2015) or VTSs, including their operation mode (Tournebize et al., 2017). However, following the concept of advective-dispersive transport (Fischer et al., 1979), the mitigating effect of dispersion on a concentration signal ~~in terms of peak concentration reduction~~ does not only depend on the magnitude of dispersion ~~imposed on that signal~~ but also on the shape of the signal. Peak concentration reduction will be stronger for a signal with a pronounced peak and low background than for a signal with a small peak and high background if both signals are exposed to the same dispersion. ~~And c~~Chemograph shapes, in turn, are dictated by processes in the contributing catchments. The influence of this chain~~is~~ of effects on ~~variability in~~ contaminant mitigation and ~~its implications for~~hence VTS efficiency~~ies~~ has~~ve~~ not been systematically investigated so far.

We hypothesize that the efficiency of contaminant mitigation in VTSs depends on the shape of the input chemographs and eventually on the factors that produce these signals in the catchment. In order to test this hypothesis we grouped chemographs of 6 contaminants mobilized in a viticultural catchment during 10 discharge events into clusters according to chemograph shape. We then compared peak concentration reduction and mass removal in a downstream VTS both among clusters and in terms of compound properties and discharge dynamics.

~~We hypothesize that mitigation efficiency in wetland systems with short residence time mainly depends on the chemograph shape of the mobilized contaminants. In order to test this hypothesis, we monitored the mobilization of 6 organic contaminants during 10 discharge events in a viticultural catchment. Then, we grouped the resulting chemographs into clusters of similar~~

~~shape. Finally, we compared R_C and R_M among clusters and assessed how they were related to other shape-related parameters, discharge dynamics, and physicochemical compound properties.In this study, we We test this hypothesis in three steps. First, we assess monitored the mobilization of six 6 organic contaminants during 10 discharge events in response to rainfall in a viticultural catchment. Second, we grouped the resulting chemographs according to shape similarity. Third, andwe evaluated the influence of the contaminant concentration signal chemograph shape on mitigation efficiencyefficiencies by comparing peak concentration reduction and mass removal rates to chemograph shape parameters as well discharge dynamics and physicochemical compound properties.~~

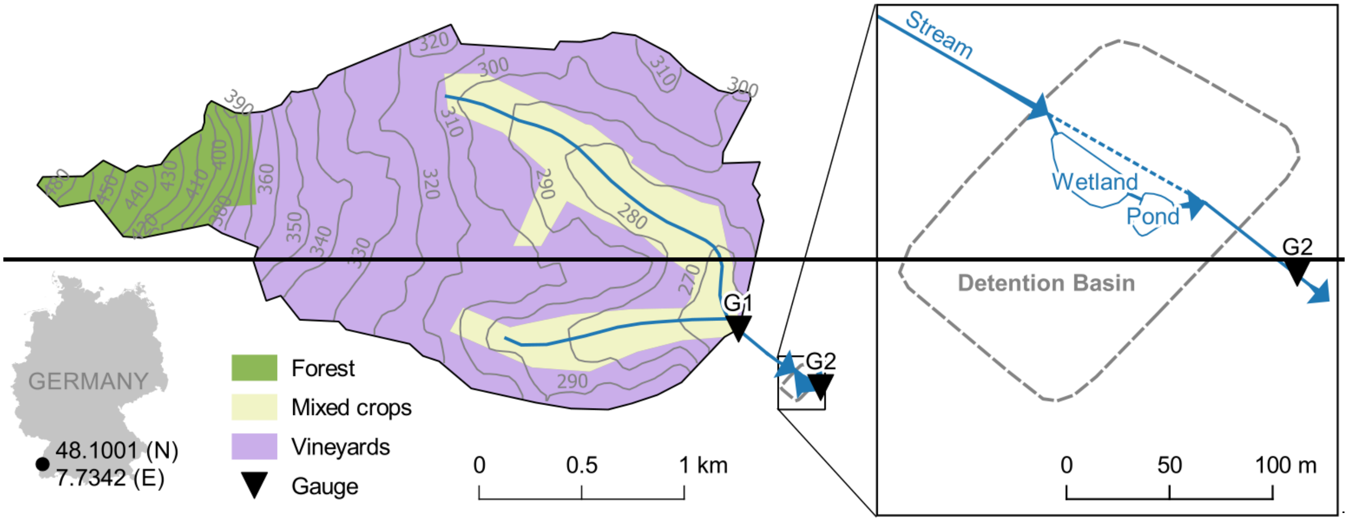
~~in a VTS located at the catchments outlet. We hypothesize that mitigation efficiency does not only depend on properties of the contaminant (e.g. sorption affinity, water and soil half-lives) or the treatment system (e.g. retention time, plant coverage), but also on the shape of the concentration signal mobilized in the catchment. To test this hypothesis, we search for patterns in contaminant flush signals by performing a cluster analysis on flow-triggered monitoring data, integrate the results into two multiple linear regression models for peak concentration reduction and mass removal, respectively, and evaluate the relative importance of the model variables.~~

2 Material and methods

2.1 Study site

The study site (Figure 1~~Fig. 1~~) is located inside a flood detention basin in ~~a~~ the 1.8 km² Loechernbach ~~headwater~~ catchment,~~,~~ southwest Germany. Catchment elevation ranges ~~from~~ between 213 m.a.s.l. at the outlet ~~to~~ and 378 m.a.s.l. in the western corner. Mean precipitation was 800 mm a^{-1} between 2009 and 2018, mean air temperature ~~was~~ 11.3 °C. Soils mainly consist of calcaric regosols which formed on ~~A~~aeolian loess and have a typical grain size distribution of 10 % sand, 80 % silt and 10 % clay. Most of ~~the land in~~ the catchm~~m~~ent ~~area~~ is dedicated to ~~viticulture on~~ large artificial vineyard terraces (71 %), while croplands occupy the ~~main~~ valley bottoms (20 %). Forest only accounts for a small portion (9 %) and is limited to the most elevated part of the catchment. This partition in land use is reflected in the main application areas of pesticide types. Fungicides are applied ~~to the~~on vineyard terraces, while ~~and~~ herbicides are mainly applied to the ~~The elevated vineyard terraces are subject to frequent fungicide application, while herbicides are applied to the~~ cropland in the flat valleys. Large parts of the catchment are drained by a ~~dense~~ sub-surface pipe network (Figure 1) ~~with a total length of about 9 km directly~~ connecting vineyards and paved roads to the main channel in the valley. T~~he~~his drainage network ~~which~~ causes fast downstream transport of storm water ~~as well asand of dissolved~~ and suspended ~~material~~ sediments (Gassmann et al., 2012). In addition, fields in the valley bottoms are drained by a secondary network of smaller and usually shorter field drains that are either connected to the primary drainage network or directly connected to the stream (Schuetz et al., 2016). A 20,000 m^3 detention basin was built at the outlet of the Loechernbach to prevent flooding of the downstream village.

Inside the detention basin, a 258 m² vegetated surface flow constructed wetland and a 105 m² retention pond (maximum depth 1.5 m) are connected in series parallel to the course of the Loechernbach stream. ~~During baseflow conditions a~~A small dam diverts all flow through the vegetated treatment systems during base flow conditions, but allows water to bybass the VTS during large discharge events. ~~In the case of large discharge events, the treatment systems are bypassed.~~ The wetland is in operation since 2010 and its succession was studied by Schuetz et al. (2012). The pond was added to the system in January 2016. ~~Both~~ The entire detention basin is sealed by an impermeable clay layer that prevents ~~infiltration~~leakage to groundwater. ~~From previous studies, it is known that w~~Water residence times~~HRT~~ ~~in the system~~ range~~s~~ from ~~about~~ less than one hour during flood events up to ~~more than a~~several days during extreme low flow conditions.



~~Fig. 1: Vegetated treatment system consisting of a vegetated stream reach, a constructed wetland and a retention pond inside a flood detention basin. Aerial picture~~

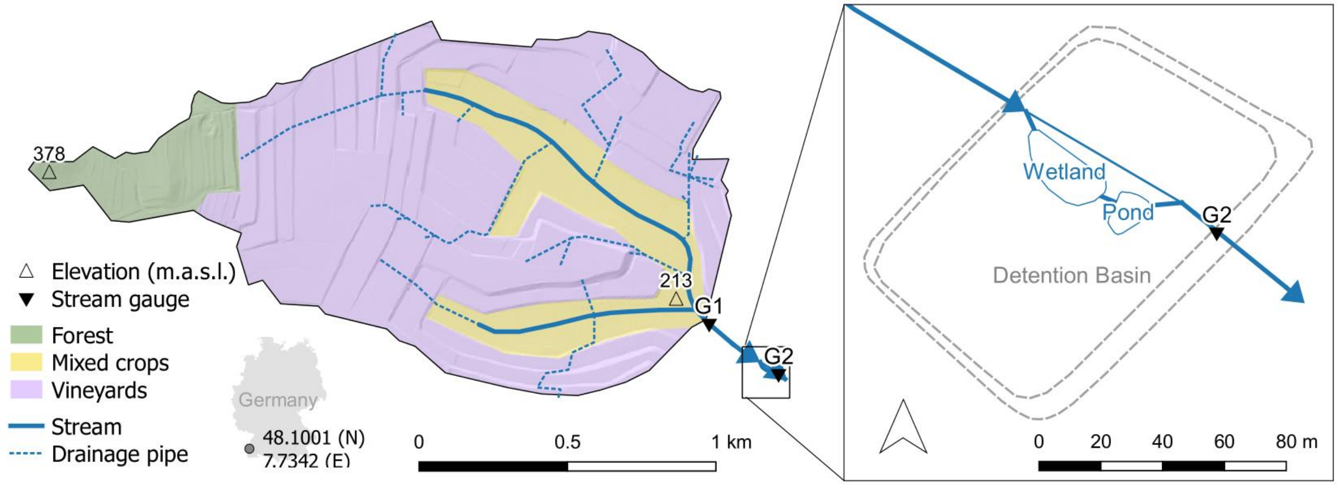


Figure 1: Loechernbach catchment and vegetated treatment system (VTS) consisting of a vegetated stream reach, a constructed wetland and a retention pond inside a flood detention basin. Shading is based on a digital terrain model with a resolution of 1 m². Location of drainage pipes is according to Gassmann et al. (2012).

2.2 Target compounds

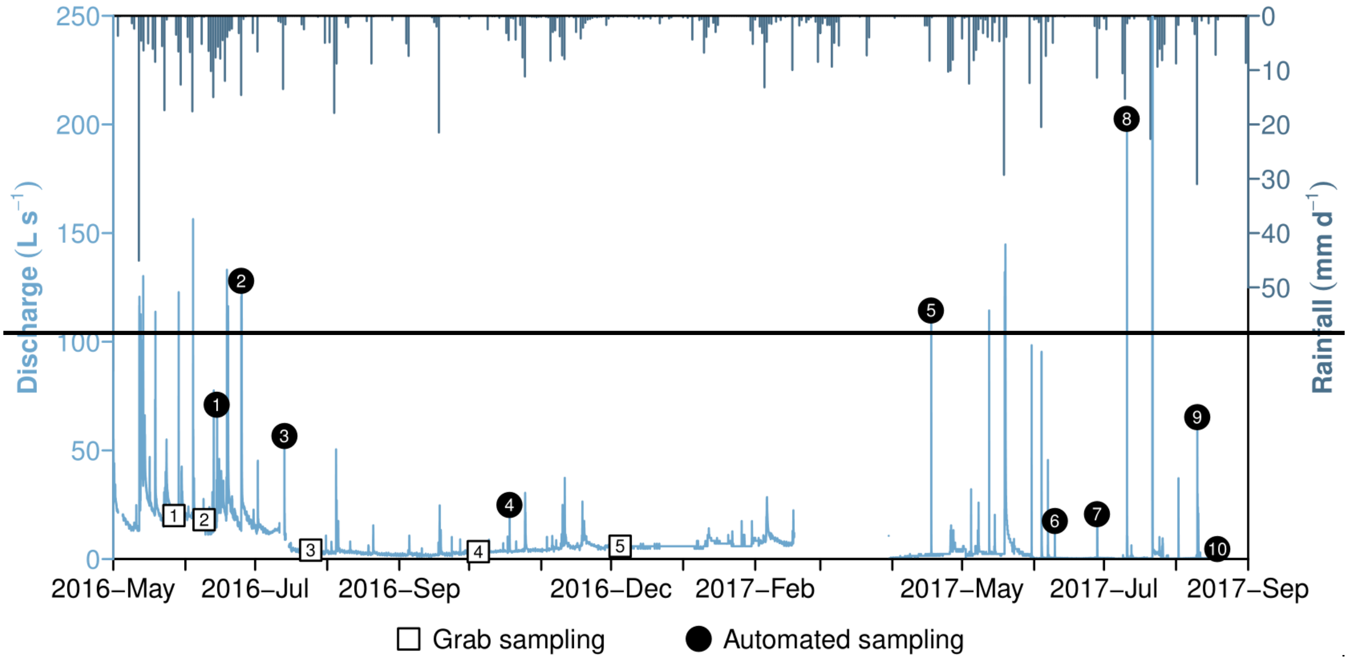
In this study, we focused on 6 target ~~The target~~ compounds ~~included~~ including the ~~two~~ fungicides boscalid and penconazole, the ~~two~~ herbicides metazachlor and flufenacet, and the ~~two~~ TPs metazachlor sulfonic acid (met-ESA) and metazachlor oxalic acid (met-OA). Selected physicochemical properties of the target compounds are listed in . According to the Pesticide Properties Data Base (Lewis et al., 2016) the contaminants can be classified as low (boscalid) to moderately soluble in water, ~~. Mobility ranges from~~ very mobile (TPs) to slightly mobile (fungicides). The target fungicides are considered moderately fast degradable in the water phase and persistent in soils, while the target herbicides are considered stable in the water phase and non-persistent in soils. TPs of metazachlor are considerably more persistent in soil than their PC. The fungicides are considered stable with respect to hydrolysis but degradable via photolysis, while the herbicides are stable regarding both.

Table 1: Physicochemical properties of the target compounds according to the Pesticide Properties Data Base (Lewis et al., 2016) including chemical formular, water solubility, organic carbon sorption coefficient (K_{OC}) as well as half lives in water (T_{50} water) and soil (T_{50} soil).

	Fungicides		***Herbicides***		***TPs***	
	Boscalid ~~(bos)~~	**Penconazole ~~(pen)~~**	**Metazachlor ~~(met)~~**	**Flufenacet ~~(flu)~~**	**~~Metazachlor sulfonic acid~~ ~~(~~met-ESA~~)~~**	**~~Metazachlor oxalic acid~~ ~~(~~met-OA~~)~~**
Chemical formula	$C_{18}H_{12}Cl_2N_2O$	$C_{13}H_{15}Cl_2N_3$	$C_{14}H_{16}ClN_3O$	$C_{14}H_{13}F_4N_3O_2S$	$C_{14}H_{17}N_3SO_4$	$C_{14}H_{15}N_3O_3$
Molecular mass (g mol^{-1})	343.2	284.2	277.8	363.3	323.4	273.3
Solubility (mg l^{-1})	4.6	73.0	450.0	51.0	-	-
K_{OC} (ml g^{-1})	772.0	2205.0	79.6	273.3	5.0	24.6
T_{50} Photolysis (d)	30.0	4.0	*stable*	*stable*	-	-
T_{50} Hydrolysis (d)	*stable*	*stable*	*stable*	*stable*	-	-
T_{50} Water (d)	5.0	2.0	216.0	54.0	-	-
T_{50} Soil (d)	246.0	117.0	10.8	19.7	123.3	90.0

2.3 ~~Monitoring setup~~ Discharge measurement and sampling procedure

Stream flow was measured every minute between April 2016 and September 2017 at two gauges about 200 m upstream of the treatment system (G1) and at its outlet (G2). Water levels at G1 were recorded inside a 1.37 m standard H-flume (Bos, 1989) by means of a pressure transducer (Decagon CTD-10) and related to discharge using a standard rating curve. At G2 water levels were measured in a rectangular cross-section 2 meters ahead of the detention basin outlet by a radar gauge (Vegapuls 61). The corresponding rating curve for G2 accounted for complete submergence of the control gate valve (Peter, 2005). Pesticide monitoring at G1 and G2 consisted of 5 manual sample collections during stationary flow conditions and 10 automated event samplings during discharge events. ~~A total of 15 sampling campaigns were performed (Fig. 2), ten of which assessed rainfall induced contaminant mobilization and five represented stationary flow. During stationary conditions, grab samples were taken at G1 and G2. Duplicates were produced by splitting the sample into 2 brown glass bottles with a volume of 1 l each.~~ Event sampling was triggered when ~~During transient flow conditions pesticide sampling was automated. When the upper gauge registered~~ a water level increase of more than 3 cm/h was registered at G1.~~,~~ A~~a~~n automatic sampler (ISCO 3700) started to fill pairs of 900 ml glass bottles at 0, 0.5, 1, 2, 6, and 12 hours after activation. A second automatic sampler (ISCO 3700) was launched at G2 following the same sampling scheme but with a time lag of one hour to account for transit between G1 and G2. All samples were recovered from the study site within 24 h after sampling and cooled until analysis. Sampling was complete except for one case. Due to accident we lost the first sample of event 9 (2017/10/08 03:30). As concentrations in the first samples were usually very low (Figure 3) and not considered to markedly influence mass calculation, we assumed that all contaminants in this sample had zero concentration and left this event in our data set.



~~Fig. 2: Sampled events during stationary (grab sampling) and transient flow conditions (automated sampling)~~

2.4 ~~Analytical methods and pesticide analysis~~Analytical methods

~~The target compounds included the two fungicides boscalid and penconazole, the two herbicides metazachlor and flufenacet, and the two TPs metazachlor sulfonic acid (met-ESA) and metazachlor oxalic acid (met-OA). Selected physicochemical properties of the target compounds are listed in Table 1. According to the Pesticide Properties Data Base~~ (Lewis et al., 2016) ~~the contaminants can be classified as low (boscalid) to moderately soluble in water. Mobility ranges from very mobile (TPs) to slightly mobile (fungicides). The fungicides are considered moderately fast degradable in the water phase and persistent in soils, while the herbicides are considered stable in the water phase and non-persistent in soils. TPs of metazachlor are considerably more persistent in soil than their PC. The fungicides are considered stable with respect to hydrolysis but degradable via photolysis, while the herbicides are stable regarding both.~~

~~Table 1: Physicochemical properties of the target compounds according to the Pesticide Properties Data Base~~ (Lewis et al., 2016) ~~including chemical formular, water solubility, organic carbon sorption coefficient (K_{OC}) as well as half lives in water (T_{50} water) and soil (T_{50} soil).~~

-	~~*Fungicides*~~		-	~~*Herbicides*~~		-	~~*TPs*~~	
-	~~**Boscalid (bos)**~~	~~**Penconazole (pen)**~~	-	~~**Metazachlor (met)**~~	~~**Flufenacet (flu)**~~	-	~~**Metazachlor sulfonic acid (met-ESA)**~~	~~**Metazachlor oxalic acid (met-OA)**~~
~~**Chemical**~~	~~$C_{18}H_{12}Cl_2N_2O$~~	~~$C_{13}H_{15}Cl_2N_3$~~		~~$C_{14}H_{16}ClN_3O$~~	~~$C_{14}H_{13}F_4N_3O_2S$~~		~~$C_{14}H_{17}N_3SO_4$~~	~~$C_{14}H_{15}N_3O_3$~~
~~**Molecular mass (g mol⁻¹)**~~	~~343.2~~	~~284.2~~		~~277.8~~	~~363.3~~		~~323.4~~	~~273.3~~
~~**Solubility (mg l⁻¹)**~~	~~4.6~~	~~73.0~~		~~450.0~~	~~51.0~~		-	-
~~**K_{OC} (ml g⁻¹)**~~	~~772.0~~	~~2205.0~~		~~79.6~~	~~273.3~~		~~5.0~~	~~24.6~~
~~**T_{50} Photolysis (d)**~~	~~30.0~~	~~4.0~~		~~*stable*~~	~~*stable*~~		-	-

~~T_{50} Hydrolysis (d)~~	~~*stable*~~	~~*stable*~~	~~*stable*~~	~~*stable*~~	-	-
~~T_{50} Water (d)~~	~~5.0~~	~~2.0~~	~~216.0~~	~~54.0~~	-	-
~~T_{50} Soil (d)~~	~~246.0~~	~~117.0~~ -	~~10.8~~	~~19.7~~ -	~~123.3~~	~~90.0~~

The following analytical methods were used for determining pesticide levels in the water samples. Analytical standards of boscalid (99.9%), penconazole (99.1%), metazachlor (99.6%), and flufenacet (99.5%) and the internal standards Diuron-D6 (99 %) and Terbutryn-D5 (98.5 %) already dissolved in acetonitrile (100 µg mL-1) were purchased from Sigma-Aldrich Chemie GmbH (Steinheim, Germany). Met-ESA (95 %) and met-OA (98.8 %) ~~and the internal standard Terbutryn-D5 (98.5 %) already dissolved in acetonitrile (100 µg mL^{-1})~~ were received from Neochema (Bodenheim, Germany). Acetonitrile (LC-MS grade; VWR International GmbH, Darmstadt, Germany) was used as organic mobile phase in chromatography and for the preparation of stock solutions. Aqueous mobile phase was prepared with ultrapure water (Membra Pure, Germany; Q1:16.6 mΩ and Q2: 18.2 mΩ.

Preparation of environmental samples (approx. 1 liter) was done by filtering with a folded filter (type 113 P Cellulose ø 240 mm). Supernatant was spiked with the internal standard Diuron-D6 (10 µl of 10 mg L-1). Extraction procedure was a solid phase extraction (SPE). Cartridges (CHROMABOND® HR-X 6 mL/200 mg) were conditioned with 10 mL methanol and washed with 10 mL pure water. ~~Samples were filtered using syringe filter units (CHROMAFIL® Xtra RC-20/25; Macharey-Nagel, GmbH & Co. KG, Germany).~~ 90 µL of the extract were spiked with 10 µl of Terbutryn-D5 as an internal standard. Each sample was a double determination. Measurements of environmental samples were conducted with a Triple Quadrupole (Agilent Technologies, 1200 Infinity LC-System and 6430 Triple Quad, Waldbronn, Germany). ~~Each sample (990 µL) was spiked with 10 µL Terbutryn-D5 as internal standard. Analysis of 5 µL of each sample was done by LC-MS/MS (Agilent Technologies, 1200 Infinity LC-System and 6430 Triple Quad, Waldbronn, Germany).~~ Mobile phases were 0.01% formic acid (A) and acetonitrile (B) with a flow of 0.4 mL min^{-1}. Gradient was as follows: 0-1 min (10% B), 1-11 min (10-50% B), 11-18 min (50-85% B), 18-21 min (85-90% B), 21-24 min (90% B), 24-26 min (90-10% B) and 26-30 (10% B). A NUCLEODUR® RP-C18 (125/2; 100-3 µm C18 ec) column (Macherey Nagel, Düren, Germany) was used as stationary phase with a set oven temperature of T = 30°C. Calibration curve were prepared in pure water. The linearity was evaluated by preparing three curves with ten calibration points in the range 1 - 500 µg/L. The standard curves were then extracted according to the protocol and analyzed using LC-MS/MS. The calculated linear regression values (R^2) were very good with R^2-values > 0.999. The linearity between peak area and concentration of substances were obtained in a range of 0 - 5 µg L^{-1}. Hence limits of detection (LOD) and quantitation (LOQ) were calculated with DINTEST (2003) according to DIN 32645 considering an enrichment factor of 5000. ~~Limits of detection (~~LOD~~)~~ and ~~quantitation (~~LOQ~~)~~ ~~were calculated with DINTEST (2003) according to DIN 32645 and~~ amounted to 0.~~35~~ 4 and 1.~~27~~ 3 ng L^{-1} (boscalid), 0.~~35~~ 3 and ~~1.29~~0.9 ng -L^{-1} (penconazole), 0.~~35~~ 3 and 1.2~~7~~ ng L^{-1} (metazachlor), 0.~~31~~ 4 and 1.3~~5~~ ng L^{-1} (flufenacet) as well as ~~1.29~~0.6 and ~~5.07~~2.2 ng L^{-1} (met-ESA) and 0.~~33~~ 5 and 1.~~23~~ 6 ng L^{-1} (met-OA) considering an enrichment factor of 5000. A detailed analysis of measurement precision can be found in the supplementary material (Text S1).

2.5 Data analysis and calculations

2.5.1 Identification of patterns in input concentration

~~2.4 Identification of patterns in input concentrations~~

Identification of patterns in input ~~concentration~~ chemographs was done by k-medoids cluster analysis - a variation of the commonly applied k-means algorithm. Both approaches partition the elements of a dataset into a predefined number k of clusters by attributing the elements to the cluster with the nearest cluster center. Optimal clustering is achieved by iteratively updating cluster centers and minimizing distance between data points and cluster centers. K-medoids differs from k-means as it uses existing points (medoids) as cluster centers instead of means and is considered more robust against extreme values and outliers (Han et al., 2012). A total of 58 concentration sequences was included in the analysis, consisting of 10 sequences per target compound, except for flufenacet which did not exceed LOQ in two events. Prior to cluster analysis, data was normalized by the maximum of each chemograph to promote that clustering represented shape, rather than differences in absolute concentration. The analysis was done using the software R (R Core Team, 2019) (version 3.6.1) using the 'pam' (partitioning around medoids) function from the 'cluster'-package (version 2.1.0) (Maechler et al., 2019). We tested clustering for k ranging between 2 and 10, the final number was determined by both visual inspection of the clusters and assessment of explanatory benefit per additional cluster (elbow method). As a result we found that k=4 resulted in the best partition.

~~by k-means cluster analysis of the contaminant input concentration sequences recorded at G1. This method was first applied in signal processing and became a popular tool in data mining where it is used for pattern recognition and also found its way into hydrology, e.g. k-means clustering has recently been applied to stream nitrate time series data~~ (Aubert and Breuer, 2016)~~. The k-means approach partitions the elements of a dataset into a predefined number k of clusters by attributing the elements to the cluster with the nearest cluster center. Optimal clustering is achieved by iteratively updating cluster centers and minimizing within cluster variance.~~

~~A total of 58 concentration sequences was included in the analysis, consisting of 10 sequences per target compound, except for flufenacet which did not exceed LOQ in two events. Prior to cluster analysis, data was normalized by the maximum of each concentration sequence to guarantee that clustering was done by signal shape, not by total concentration. The analysis was done in the software R~~ (R Core Team, 2019)~~ using an algorithm by~~ Hartigan and Wong (1979)~~. We tested clustering for k ranging between 2 and 10, the final number was determined by both visual inspection of the clusters and assessment of explanatory benefit per additional cluster (elbow method). We found that k=4 resulted in the best partition.~~

~~In order to integrate the results of the clustering into further analysis, we developed the dispersion sensitivity index i_{DS}. Following the concept of advective-dispersive solute transport, relative peak concentration reduction can be expected to be higher for concentration signals with clearly defined peaks than for those with relatively flat peaks compared to the background. The dispersion sensitivity index was correspondingly defined as the relative portion of peak concentration ($\hat{C}_{in}$) susceptible to dispersion according to Eq. 1:~~

$$\cancel{i_{DS} = \left(1 - \frac{C_{n,in}}{\hat{C}_{in}}\right)^{2},} \qquad \cancel{(1)}$$

~~where $C_{n,in}$ is the concentration in the last sample collected, considered as a reference for post-peak concentration.~~

2.5.2 Contaminant mitigation

~~2.5 Contaminant retention~~

Contaminant retention was assessed in terms of both peak-concentration reduction rate (R_C) and mass removal rate (R_M). R_C was calculated in accordance with other studies (Eq. 2), e.g. Elsaesser et al. (2011), Stehle et al. (2011) and Passeport et al. (2013):

$$R_C = \frac{C_{in,max} - C_{out,max}}{C_{in,max}} \cdot 100\,\%, \cancel{R_C = \frac{\hat{C}_{in} - \hat{C}_{out}}{\hat{C}_{in}} \cdot 100\,\%,} \tag{2}$$

where ~~$\hat{C}_{in}$~~ $C_{in,max}$ and ~~$\hat{C}_{out}$~~ $C_{out,max}$ are peak concentrations registered at the inlet and outlet sampling points, respectively. R_M was calculated analogously from the input (M_{in}) and output contaminant mass (M_{out})~~ as shown in Eq. 3~~:

$$R_M = \frac{M_{in} - M_{out}}{M_{in}} \cdot 100\ \%\,, \qquad (3)$$

Contaminant masses were calculated from discharge at G1 and G2 and linearly interpolated contaminant concentrations. ~~For comparability with automated sampling, contaminant masses during stationary flow were referred to a period of 12 h during which constant concentrations and flow were assumed.~~ As water level data from G2 showed evidence for inaccuracy during low flows ~~due~~ as result ~~to~~of the ~~constant~~ rectangular shape of the measuring~~-~~ cross-section at G2, we ~~assumed that the VTS was in equilibrium during stationary flow conditions and used flow from G1 for calculation of both M_{in} and M_{out}.~~ did not assess mass removal during stationary flow conditions. As we did not sample the wetland sediments or plants, the mass removal rate calculated following the above procedure describes ~~Following this procedure, mass removal assessment is limited to the liquid phase as contaminants adsorbed to sediments are not taken into account.~~ the relative difference of dissolved contaminant mass entering and leaving the wetland within the duration of the sampling procedure. It is therefore not independent of the wetland's water balance:

$$W_B = \frac{Q_{in,mean} - Q_{out,mean}}{Q_{in,mean}} \cdot 100\ \%\,, \qquad (4)$$

where $Q_{in,mean}$ and $Q_{out,mean}$ are the discharge at G1 and G2, respectively, averaged over the duration of the sampling procedure at both gauges. W_B was positive, if more water entered the wetland than left the wetland during the sampling procedure, and negative in the opposite case.

2.5.3 Dispersion sensitivity of chemographs

We defined a dispersion sensitivity index as follows:

$$i_{DS} = \frac{C_{in,max} - C_{in,n}}{C_{in,max}}\,, \qquad (5)$$

where $C_{in,n}$ is the concentration in the last sample and $C_{in,max}$ is the peak concentration of a chemograph recorded at the inlet of the VTS (G1). In other words, i_{DS} represents the fraction of the concentration peak that can potentially be flattened by dispersion.

~~2.6 Identification of influential variables~~

~~In order to identify influential variables for contaminant mitigation we constructed two separate multiple linear regression models for R_C and R_M and attributed relative importances to the model variables. As TP peak concentration data turned out to be affected by low sampling frequency during late stages of the sampling procedure, we decided to focus on PCs and exclude TPs. This decision reduced the number of data points but prevented the model from being affected by the high variability in TP mitigation efficiencies resulting from imperfect sample coverage. Model construction was done as follows. First, an initial set of potential model variables was selected based on literature and practical considerations for each model. Then all-subsets regression~~ (Lumley, 2017)~~ was performed on the initial set of variables, i.e. all possible numbers of variables and variable combinations were tested and evaluated in terms of explanatory power as expressed by adjusted R².~~

~~The initial variable sets of the two models are represented by the light columns in the background of Fig. 6. They included physicochemical contaminant properties such as the organic carbon adsorption coefficient (K_{OC}) and water solubility (Solub). In contrast to other studies~~ (Bundschuh et al., 2016; Stehle et al., 2011)~~, we decided to not include the degradation half-lives listed in Table 1 as even the lowest of these values (Water T_{50} of penconazole is reported to be 2 days) would not have produced measurable effects in the short retention time of approx. one hour during event conditions. Hydraulic conditions were represented by mean discharge ($\bar{Q}_{in}$) and mean hydraulic retention time (HRT). Variables considered relevant for only one of the models were input peak concentration ($\hat{C}_{in}$) and dispersion sensitivity (i_{DS}) in the R_C model as well as input mass (M_{in})~~

~~and the water balance error e_W which was included in the R_M model. A variable identified as important by other studies~~ (Stehle et al., 2011; Bundschuh et al., 2016; Elsaesser et al., 2011) ~~but not included in our study is vegetation coverage. VTS properties such as vegetation coverage may be essential for comparison of different systems~~ (Lange et al., 2011; Stang et al., 2014) ~~or states of systems~~ (Schuetz et al., 2012)~~. However, this was not the main purpose of our study.The data set was checked for multi-collinearity of independent variables by calculating Pearson correlation coefficients and variance inflation factors (VIF). Strongest correlation was found between $\bar{Q}_{in}$ and HRT (R=0.85). VIF for these variables were between 3 and 5. This is below the standard threshold of 10, but as~~ Hair (2010) ~~points out that in models with limited data, comparatively small VIF values may indicate substantial collinearity, we decided to keep the variables that was stronger correlated with the mitigation rates ($\bar{Q}_{in}$) and exclude the other one (HRT) from the model building routine. The exclusion of HRT, however, only marginally changed the model output.~~

~~Data points were considered outliers and removed from the data set, if at least one of the following condition was met: (1) Cook's distance exceeded one, indicating particularly influential points, or (2) R_C or R_M, in the respective models, deviated from the mean by more than 2 standard deviations. The number of observations in the data sets passed to all-subset regression was 36 for the R_C model and 37 for the R_M model. Relative importance of the predictors in the resulting model was determined by decomposing total R² into non-negative contributions based on the corresponding sum of squares~~ (Grömping, 2006)~~. As attribution of explanatory power to predictors can depend on the order in which they are added to the model, we applied the LMG measure~~ (Lindemann et al., 1980) ~~which averaged predictor importance over all possible orderings.~~

3 Results

3.1 ~~Input concentration and mass~~Contaminant mobilization

Contaminant concentrations in stream water (G1) differed clearly depending on the flow conditions (Figure 2). ~~Input concentrations (G1) differed clearly under the different flow conditions.~~ During stationary flow, concentrations of boscalid and the TPs of metazachlor ranged in the order of tens of nanograms~~were detected in almost all samples~~, while penconazole, metazachlor and flufenacet only occasionally exceeded the LQ~~were only found occasionally. Concentrations were usually in the order of tens of nanograms per litre (Fig. 2).~~. During ~~transient flow~~discharge events in contrast, peak concentrations varied from a few nanograms (flufenacet) to several ~~miligrams~~milligrams per liter (boscalid) spanning a range of 6 orders of magnitude. ~~all contaminants were found in all samples. Concentrations were generally much higher than during stationary flow and the concentration difference depended on the contaminant.~~ Concentration increase during events compared to stationary flow was different among the compounds. Median concentration of boscalid increased by a factor of 48, while concentrations of met-ESA and met-OA only increased by a factor of 3 and 5, respectively. Similar patterns were found for contaminant mass. Contaminant mass mobilized in the catchment during discharge events ranged from several hundreds of micrograms (flufenacet) to several hundreds of milligrams (boscalid) and even several grams in exceptional cases (boscalid, metazachlor). Based on compound medians, about 76 times more boscalid but only about 4 times more met-OA were transported during discharge events than during an equally long period under stationary flow conditions.

~~Medians of penconazole, metazachlor and flufenacet were not calculated due to lack of data during stationary flow. The increase in concentrations translated into an even stronger increase in input load. During stationary flow, input mass related to a 12 hour period usually was in the order of tens of micrograms. During transient flow, input mass increased by about 4 orders of magnitude in the case of boscalid and about 2 orders of magnitude in the cases of met-ESA and met-OA. In extreme cases, several grams of boscalid (and metazachlor in one exceptional case) entered the treatment system during a single event.~~

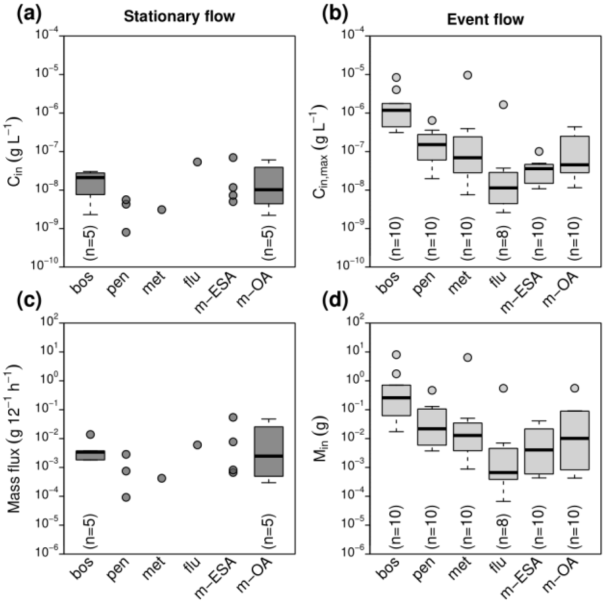


Figure 2: Peak concentrations at G1 during stationary flow (a) and flow events (b) as well as contaminant mass flux during stationary flow (c) and transported mass per event (d). Boxplots indicate median and interquartile range (IQR). Whiskers indicate extreme points within 1.5 times the IQR from the boxes, circles indicate points outside this range.

~~**Fig. 2: Peak concentrations and input mass at G1 of the contaminants boscalid (bos), penconazole (penconazole), metazachlor (met), flufenacet (flu), metazachlor sulfonic acid (met-ESA) and metazachlor oxalic acid (met-OA) during stationary and transient flow conditions. Boxplots are not shown if less than five data points were available for the corresponding contaminant. Input mass during stationary conditions was referred to a period of 12 hours.**~~

The 10 events were characterized by different discharge magnitudes and dynamics (Figure 3). Mean discharge during the events ranged between 0.7 (E10) and 32.0 L s^{-1} (E2) with respective peak values between 4.4 (E10) and 199.7 L s^{-1} (E2). The recorded event hydrographs included events with one single discharge peak (E4, E5, E6, E10), with one major peak followed by one or more secondary peaks (E2, E3, E7, E9), and events in which a major peak followed an earlier smaller peak (E1, E8). In most cases discharge had recessed to pre-event levels by the end of the 12-hour sampling procedure, only E1 and E2 showed ongoing flow recession. In many cases, concentrations in the final event samples were still elevated compared to pre-event conditions. However, due to flow recession, mass flux was usually very low by the time the last sample was collected (Figure S1).

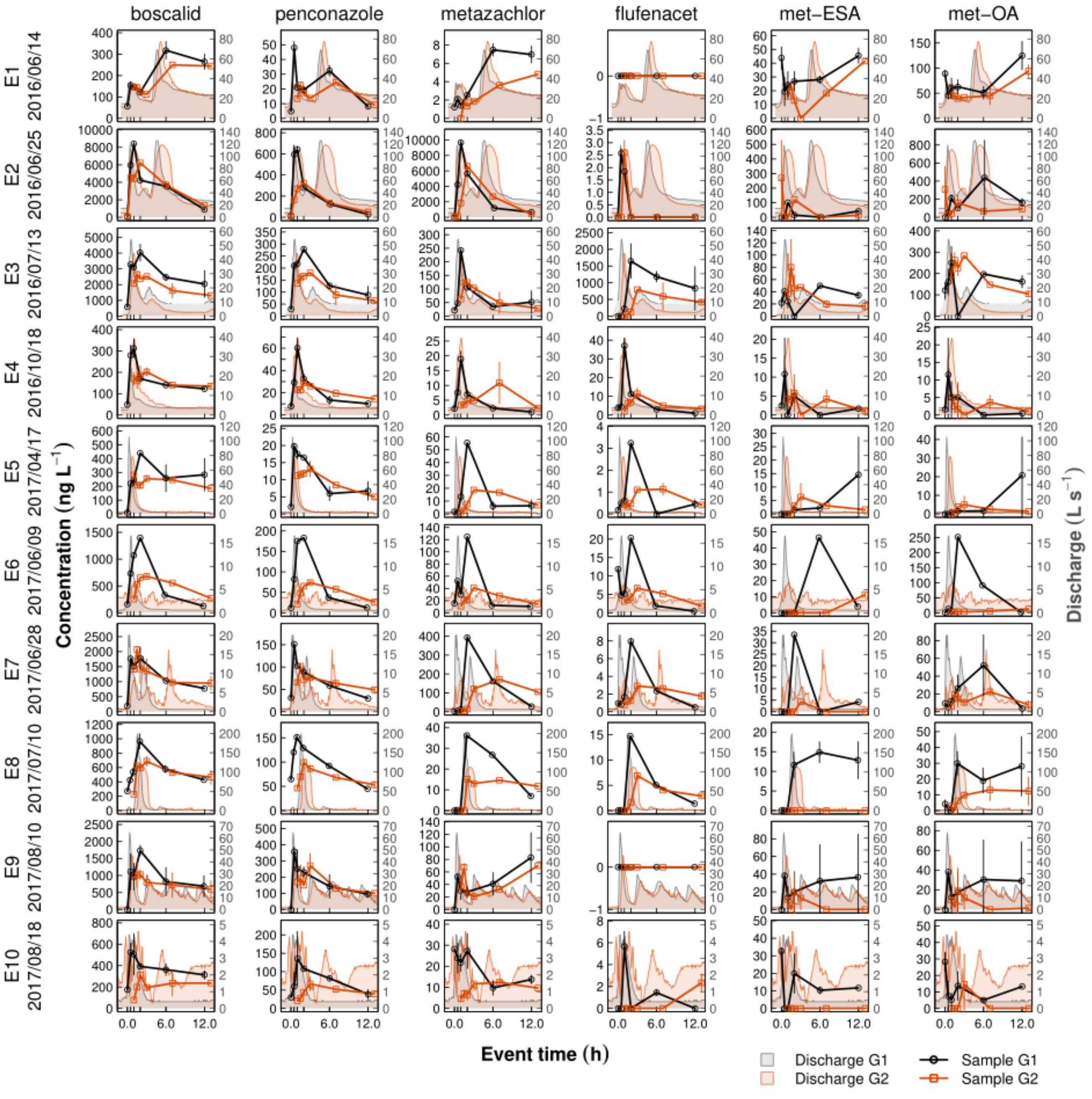


Figure 3: Contaminant concentration at the inlet gauge G1 and the outlet gauge G2 of the 6 target compounds during 10 discharge events. Data points represent means of duplicate samples and standard deviation (error bars).

~~4 Results and discussion~~

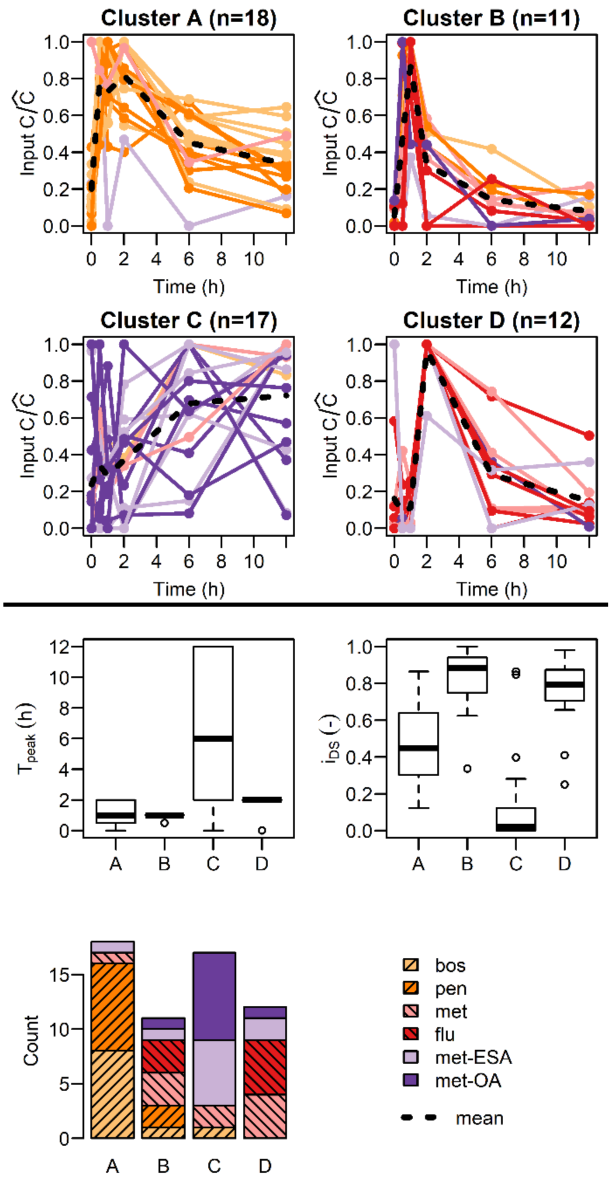
~~4.1~~3.2 ~~Contaminant mobilization patterns~~Patterns in chemographs

Cluster A (Figure 4) was characterized by absence of a clear peak during the first two hours of sampling but elevated concentrations during later times, resulting in low i_{DS}. Cluster B showed a quick response, i.e. concentrations increased sharply within the first 30 minutes. Concentrations were the highest of all clusters and still elevated in the last sample compared to pre-event levels. Cluster C was characterized by a clear peak within the first two hours and a low tailing and was the cluster with highest median i_{DS}. Cluster D showed the most inconsistent pattern and maximum concentrations appeared later compared to clusters B and C. A relatively clear pattern was evident in the attribution of compounds to the clusters. Chemographs of the fungicides boscalid and penconalzole were mainly assigned to cluster B, while the herbicides and the TPs were assigned to the remaining three clusters. Cluster A was composed of herbicide and TP chemographs, particularly from events with multiple discharge peaks. Cluster D represented chemographs of herbicides and TPs mainly during the events E5 to E8 which were all characterized by sharp discharge peaks during periods of generally low flow (Figure 3). Almost all chemographs of the events E2 and E4 were attributed to cluster C.

~~mainly represented the fungicides (boscalid and penconazole) was characterized by a short median peak arrival time (T_{peak}) of 1 h and an elevated shoulder, i.e. concentration often had not dropped back to zero by the end of the sampling procedure. Due to the pronounced tailing, the dispersion sensitivity index reached intermediate values. Cluster B was of mixed composition and showed a similarly quick increase (T_{peak}=1 h). Cluster B differed from cluster A by a lower tailing and corresponding higher dispersion sensitivity. Cluster C was dominated by the TPs of metazachlor and was overall less concise but clearly distinguishable from the other clusters by the late peak (T_{peak}=6 h). In cluster C peak and late-time concentration were often similar which resulted in very low dispersion sensitivity. Cluster D consisted mainly of the herbicides (metazachlor and flufenacet) and was similar to cluster B but peak concentration was reached about one hour later (T_{peak}=2 h). With the exception of cluster B which rather represented similar events (event 1 and event 4 in Fig. 2), overall clustering was controlled by similar behavior of contaminant groups.~~

~~We interpret these findings as an indication of different source areas and input transport pathways. In the study catchment fungicides (cluster A) are in fact mainly applied to the elevated vineyard terraces and likely to quickly reach the stream via surface runoff and the drainage network. Herbicides (cluster D), in contrast, are rather applied to croplands in the flat valley areas and transported to the stream more slowly due to lower terrain slope. Rising water levels in the stream may further reduce the hydraulic gradient in drainage pipes and thereby reduce flow velocity. The TPs (cluster C) have the same geographical source area as their PC but are formed in the soil and seem to be transported comparatively slowly to the drainage network via matrix leaching. This interpretation is in line with findings from a field study by~~ Doppler et al. (2012) ~~who showed that a combination of macropores and tile drains may act as a quick export pathway of pesticides to the stream. Slower export pathways of TPs compared to PCs were also reported by~~ Gassmann et al. (2013) ~~who found that soil matrix leaching rates to tile drains played a bigger role for TPs than for PCs in a Swiss catchment. This finding was attributed to higher mobility and higher amounts of initial residues of TPs compared to PCs in the soil matrix. These explications also seem plausible in the catchment investigated in this study.~~

~~One ambiguous aspect to k-means cluster analysis is the selection of the number of clusters. While we found k=4 to result in interpretable clusters, other numbers of clusters would be equally justifiable based on explanatory power gained per additional cluster. In order to make this somewhat arbitrary decision more transparent, we also checked changes in cluster composition if three or five clusters were allowed. We found that in both cases the fungicides and TPs were still largely separated, while the remaining clusters represented intermediate properties, i.e. the overall interpretation of the clusters would not change. In the present study the function of the cluster analysis was mainly to illustrate differences in contaminant behavior, however, our results indicate a potential for its application for data exploration in catchments where less prior knowledge on catchment hydrology is available.~~



~~**Fig. 3: Clustered input chemographs, peak arrival time (T_{peak}), dispersion sensitivity index (i_{DS}) and cluster composition. Dashed lines indicate cluster means.**~~

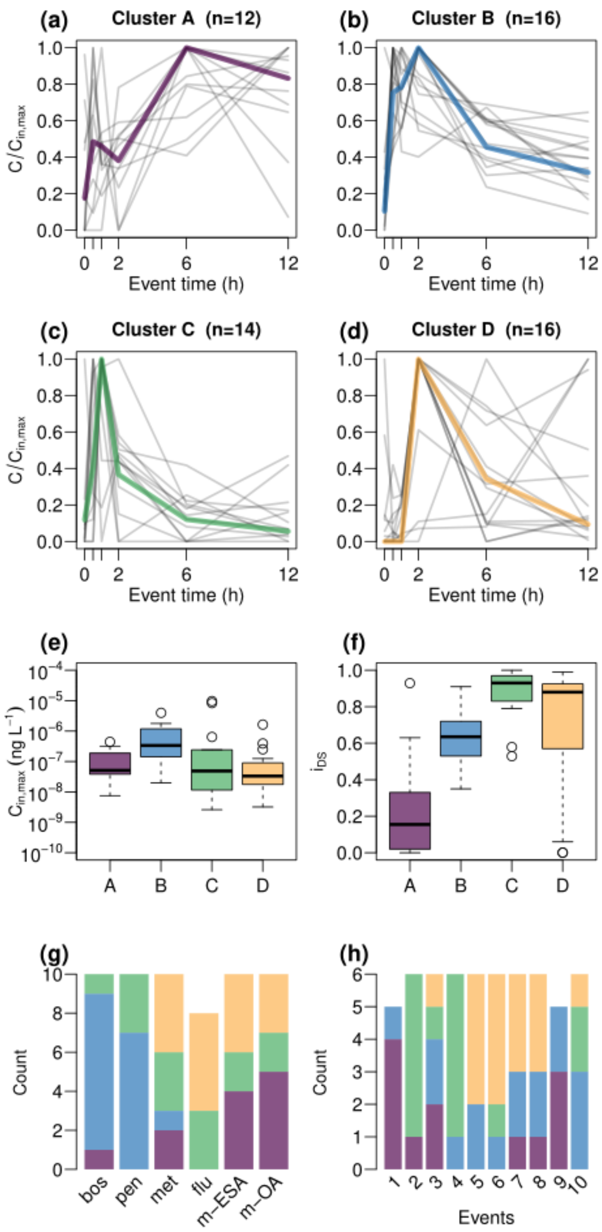


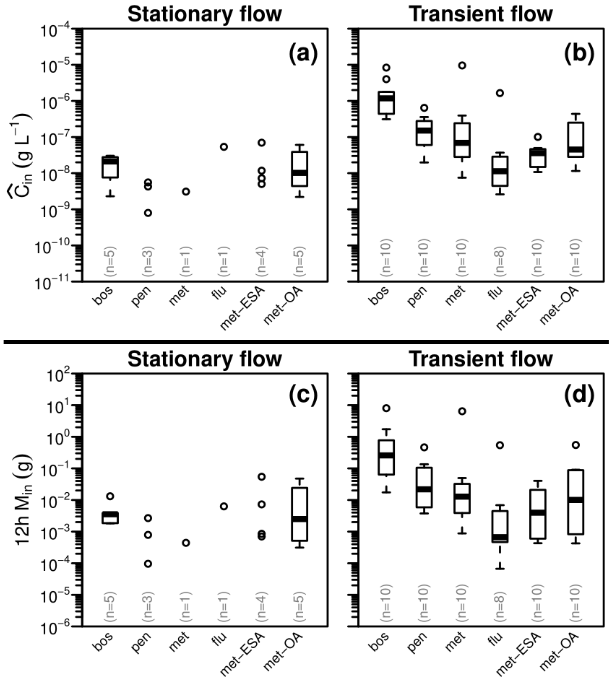
Figure 4: Clustered event chemographs as well as maximum input concentration, dispersion sensitivity index, and attribution of compounds and events to the different clusters

~~4.2 Input concentrations and loads~~

~~Input concentrations (G1) differed clearly under the different flow conditions. During stationary flow, boscalid and the TPs of metazachlor were detected in almost all samples, while penconazole, metazachlor and flufenacet were only found occasionally. Concentrations were usually in the order of tens of nanograms per litre (Fig. 4). During transient flow, all contaminants were found in all samples. Concentrations were generally much higher than during stationary flow and the concentration difference depended on the contaminant. Compared to stationary flow, median concentrations of boscalid increased by a factor of 48,~~

~~while concentrations of met-ESA and met-OA only increased by a factor of 3 and 5, respectively. Medians of penconazole, metazachlor and flufenacet were not calculated due to lack of data during stationary flow. The increase in concentrations translated into an even stronger increase in input load. During stationary flow, input mass related to a 12 hour period usually was in the order of tens of micrograms. During transient flow, input mass increased by about 4 orders of magnitude in the case of boscalid and about 2 orders of magnitude in the cases of met-ESA and met-OA. In extreme cases, several grams of boscalid (and metazachlor in one exceptional case) entered the treatment system during a single event.~~

~~Concentrations detectable at the catchments may depend on many factors including stream size~~ (Lorenz et al., 2017)~~ and land-use which complicates direct comparison. Peak concentrations of fungicides detected in this study, however, were in the range of those detected by~~ Bundschuh et al. (2016)~~ in different streams draining viticultural catchments in Southwest Germany. The fact that highest peak concentrations were detected for the fungicides reflects both the land-use distribution in the catchment (higher percentage of vineyards compared to cropland) and the different application practices (higher application frequencies of fungicides). The increase in PC to TP concentration ratio from stationary to transient flow conditions indicates that TPs are exported via a permanently active pathway such as groundwater. In contrast, PCs are rather transported via periodically active pathways that respond quickly to rainfall. This interpretation is in line with the patterns identified in section 3.1 and findings of~~ Gassmann et al. (2013)~~.~~



~~Fig. 4: Peak concentrations and input mass at G1 of the contaminants boscalid (bos), penconazole (penconazole), metazachlor (met), flufenacet (flu), metazachlor sulfonic acid (met-ESA) and metazachlor oxalic acid (met-OA) during stationary and transient flow conditions. Boxplots are not shown if less than five data points were available for the corresponding contaminant. Input mass during stationary conditions was referred to a period of 12 hours.~~

3.3 Contaminant mitigation in the wetland ~~Contaminant retention inversely related to input concentration and load~~

Contaminants were mitigated in the wetland in terms of both R_C (Eq.1) and R_M (Eq.2). R_C (Figure 5a) was close to zero for boscalid and poorly constrained for the remaining compounds during stationary conditions, partly due to insufficient number of detections. During discharge events, in contrast, peak concentrations of all compounds were clearly reduced. While R_C was narrowly constrained for the fungicides and herbicides, TPs exhibited higher variability. Mean R_C and corresponding standard deviations were 29.8 ± 18.4 % (boscalid), 42.1 ± 11.5 % (penconazole), 47.9 ± 16.4 % (metazachlor), 53.8 ± 22.6 % (flufenacet), 29.5 ± 84.7 % (met-ESA), and 47.9 ± 29.5 % (met-OA), respectively. R_C was clearly different among chemograph clusters with lowest values in cluster A and highest values in cluster D (Figure 5b). Moreover, R_C was higher for the lower half of peak concentrations than for the upper half (Figure 5c) and systematically increased with dispersion sensitivity (Figure 5d). R_C was also related to discharge conditions. Highest R_C values were reached, when mean discharge was low (Figure 5e) but the ratio of maximum to mean discharge was elevated (Figure 5f), i.e. in events characterized by low pre-event discharge and sharp discharge peaks (in particular events attributed to cluster D in Figure 4). Although there was evidence for major water surpluses and deficits in the event water balance between G1 and G2, particularly in events with low discharge such as events E6 and E10, an imbalanced water balance had only minor effects on R_C (Figure 5g). We did not find clear relationships between R_C and compound properties such as Kf_{OC}, water solubility or soil half-live (Figure 5h-j).

Relative mass removal during discharge events (Figure 6a) resulted in smaller rates and higher variability compared to R_C. Mean R_M and corresponding standard deviations were 7.7 ± 29.6 % (boscalid), 17.3 ± 26.0 % (penconazole), 18.1 ± 27.8 % (metazachlor), 27.0 ± 28.1 % (flufenacet), 35.2 ± 68.4 % (met-ESA), and 44.0 ± 28.7 % (met-OA). These values show that mass removal was limited for most compounds. Although the general pattern in R_M for the different compounds was similar to R_C, behavior of R_M among the chemograph clusters was different. While R_C increased from cluster A to cluster B and C, R_M decreased (Figure 6b). Cluster D exhibited high values of both R_C and R_M. No clear response was found to different levels of input mass (Figure 6c), however, median R_M was lowest when peaks in chemographs were sharpest (Figure 6d). This means the relationship of R_M to increasing sharpness of chemograph peaks was inverted compared to R_C. R_M was not obviously related to discharge dynamics, neither to mean discharge (Figure 6e), nor to the ratio of maximum to mean discharge (Figure 6f). Disregarding events with very low discharge (events E6 and E10), it seemed possible that much of R_M was the result of water imbalances during the events (Figure 6g). However, R_M of most chemographs plotted above the 1:1 line of R_M and relative water balance, indicating that R_M was higher than water imbalance would explain. R_M showed a tendency to decrease with increasing Kf_{OC} (Figure 6h), but no clear pattern was found for solubility (Figure 6i) and soil half-live (Figure 6j).

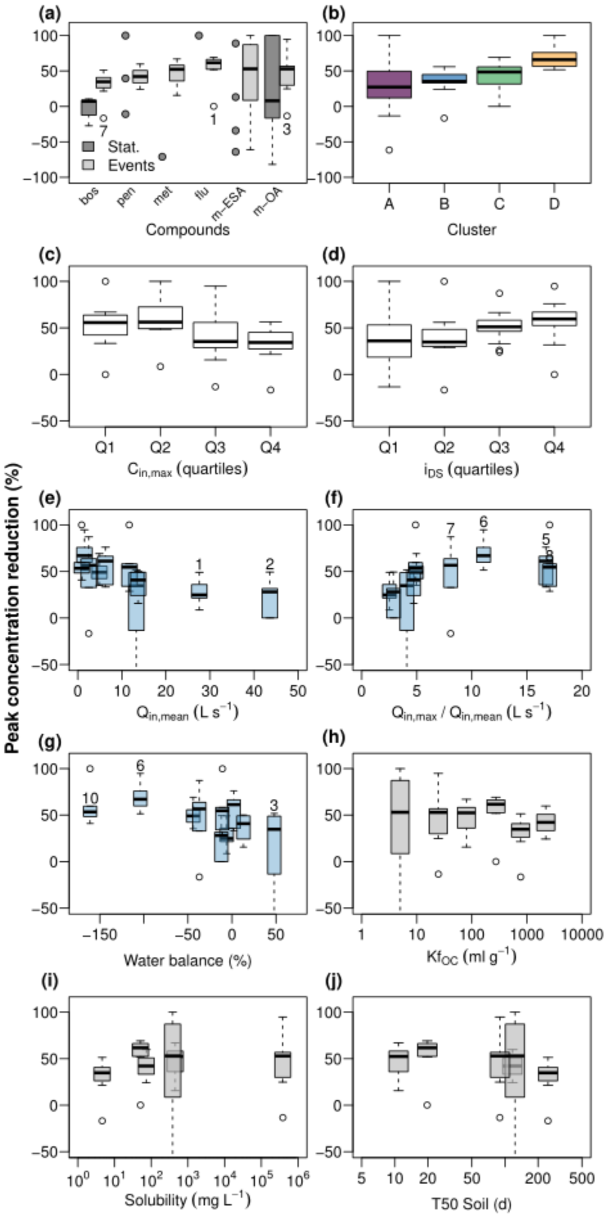


Figure 5 Contaminant peak concentration reduction in the wetland (a) during stationary and event flow conditions and its relationship to chemograph properties, discharge conditions, and physiochemical properties of the target compounds. Chemograph properties include clustering (b), peak concentrations (c), and ratio of concentrations during the peak and in the tailing (d). Discharge conditions include mean discharge at G1 (e), ratio of maximum to mean discharge at G1 (f), and water balance between G1 and G2 (g). Event numbers are shown for selected events. Compound properties include organic carbon sorption coefficient (h), solubility in water (i) and soil half-live (j).

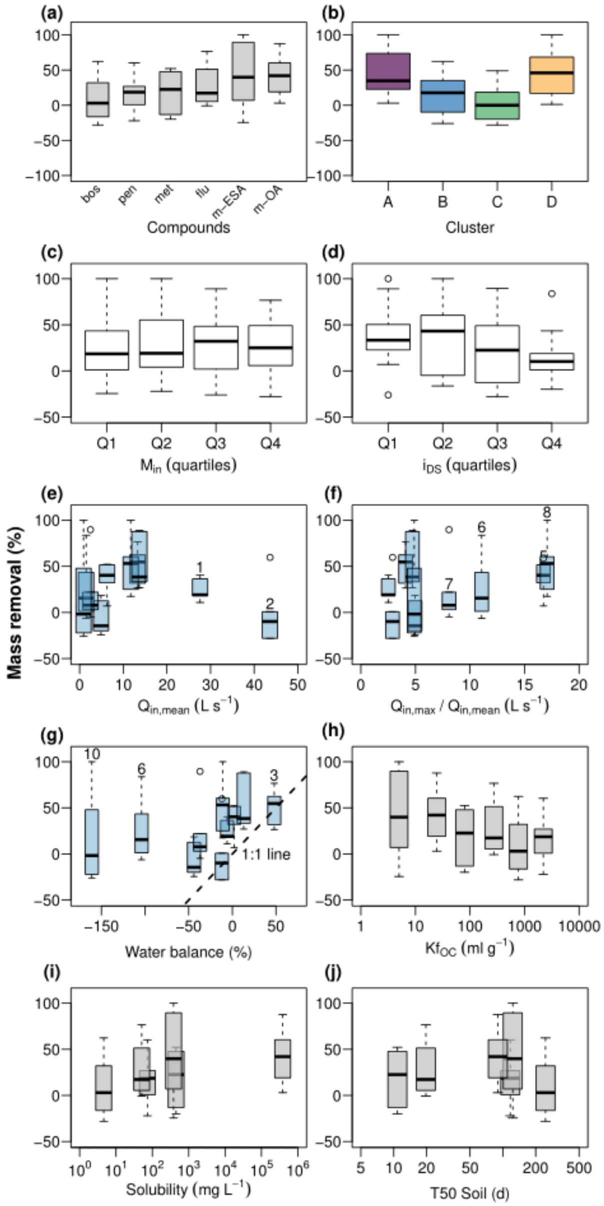
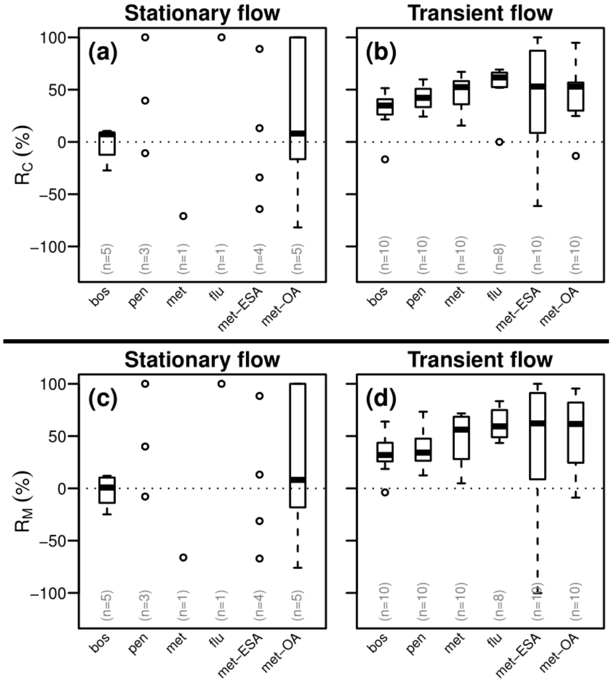


Figure 6: Contaminant mass removal in the wetland (a) during event flow conditions and its relationship to chemograph properties, discharge conditions, and physiochemical properties of the target compounds. Chemograph properties include clustering (b), peak concentrations (c), and ratio of concentrations during the peak and in the tailing (d). Discharge conditions include mean discharge at G1 (e), ratio of maximum to mean discharge at G1 (f), and water balance between G1 and G2 (g). Event numbers are shown for selected events. Compound properties include organic carbon sorption coefficient (h), solubility in water (i) and soil half-live (j).

~~Different behavior in terms of contaminant retention was observed depending on flow conditions (Fig. 5). When flow was stationary, R_C (Eq.1) and R_M (Eq.2) of boscalid were close to zero. Mitigation rates of the remaining compounds were poorly constrained (met-OA) or based on an insufficient number of detections (penconazole, metazachlor, flufenacet, met-ESA) for identification of a clear pattern. When, in contrast, flow was transient, median R_C of the individual compounds ranged from 34.8 (boscalid) to 61.6 percent (flufenacet) and median R_M ranged from 31.9 (boscalid) to 61.2 percent (met-ESA). In a few cases, calculation of R_C (n=4) and R_M (n=4) resulted in negative values. For both R_C and R_M variability was higher for TPs than PCs.~~

~~The finding that R_M was similar to R_C is surprising at first sight because degradation of contaminants usually happens on larger time scales in relation to HRT in the studied VTS or comparable systems~~ (Elsaesser et al., 2011)~~. Assessment of mass loss in our study, however, was limited to the liquid phase. Contaminant mass loss could therefore be the result of e.g. incorporation of contaminants by plants or adsorption to organic matter inside the VTS~~ (Stang et al., 2014) ~~or temporal trapping of highly contaminated portions of flow. Temporal trapping may have occurred as water flushed into remote areas of the detention basin where it was retained for durations exceeding the end of our event sampling scheme. Small, disconnected depressions filled with water were in fact observed in the field after flood events. If such areas are temporally flooded, degradation conditions may be similar to those in VTS operated in batch mode, i.e. alternating oxic-anoxic conditions enhanced degradation as found by~~ Maillard et al. (2016)~~. Considering the different mobilization behavior of the contaminants, mass removal due to trapping may be different for the contaminants depending on which portion of flow was intercepted.~~

~~Experimental uncertainties may be associated to different discharge measurement methods at the two gauges and resulting underestimation of contaminant mass. This in turn, may have caused overestimation of mass loss. However, relative water balance errors were usually smaller than R_M. In order to check for a systematic relationship of water balance and RM, we included water balance as a potentially influential variable into the regression model building procedure. The sampling scheme seemed not to be the major source of error for PC mass, as PC concentrations had usually recessed to background levels when sampling ended (Fig. 3). This, again, was different for the TPs that arrived later and whose mass balances were therefore not equally well captured by the sampling scheme.~~



~~Fig. 5: Peak concentration reduction and mass removal rates of the contaminants boscalid (bos), penconazole (penconazole), metazachlor (met), flufenacet (flu), metazachlor sulfonic acid (met-ESA) and metazachlor oxalic acid (met-OA) during stationary and transient flow conditions. Numbers in brackets indicate the number of data points. Boxplots are not shown if less than five data points were available for the corresponding contaminant.~~

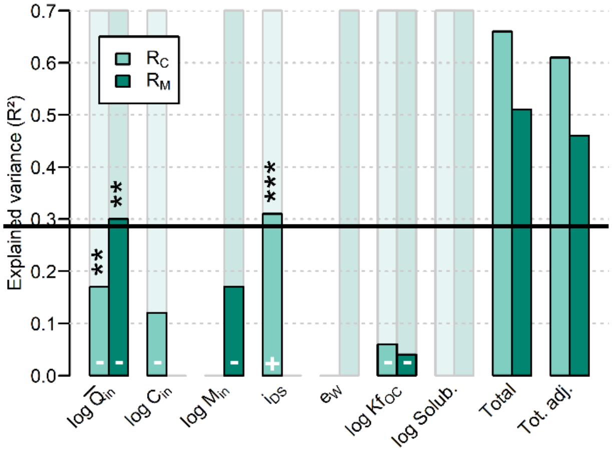
~~4.3 Dispersion sensitivity dominates peak concentration reduction~~

~~The linear regression models explained 66 and 51 percent of variance in R_C and R_M, respectively (Fig. 6). This means that the models generally described concentration reduction more accurately than mass removal. The models also differed in terms of importance attributed to the variables. In terms of relative variable importance, the R_C model of parent compounds was dominated by the effects of i_{DS} which accounted for about half of the explained variance. A secondary significant variable was discharge, contributions from input concentration and K_{OC} were not significant. Mass removal was best described by a combination of discharge and input mass. The latter, however, was not significant, nor was K_{OC}. The water balance error was excluded by the all subset regression model building procedure because it only marginally improved explanatory power of the R_M model.~~

~~According to the R_C model, R_C was positively related to i_{DS}, and negatively to mean discharge and input concentration. These results may be considered physically reasonable as high peak concentrations are likely to coincide with pronounced tailings and therefore low dispersion sensitivity. The importance of discharge in our case is considered equivalent to the importance of HRT found in other studies~~ (Bundschuh et al., 2016; Stehle et al., 2011) ~~as these variables were subject to strong negative correlation and are physically related. We therefore regard these results as a confirmation of our hypothesis that the input signal shape should be relevant for R_C.~~

~~In contrast to other studies, we did not find a significant relationship of R_C with K_{OC}~~ (Stehle et al., 2011) ~~or solubility~~ (Bundschuh et al., 2016)~~. In fact, the small contribution to explanatory power of K_{OC} is based on a negative relationship, which contradicts the general opinion in literature~~ (Vymazal and Březinová, 2015) ~~and is probably an artifact resulting from the low~~

~~number (n=4) of compounds and thus different K_{OC} values included in the model. While i_{DS} clearly improved the R_C model, no such specific variable was found for the R_M model. The lower explanatory power of the R_M model indicates that some important explanatory variable may be missing, possibly because temporal trapping of flow portions was not represented by the selected model variables.~~



~~Fig. 6: Overall explanatory power of R_C and R_M models (total R^2 and total adjusted R^2) and contributions of model variables in terms of explained variance. Light columns in the background indicate that the variable was permitted for the model building procedure. Asterisks above columns indicate significance levels of the model variables according to t-statistics ($p<0.05$: '*', $p<0.01$: '', $p<0.001$: '***'). The white '+' and '-' symbols indicate whether the underlying relationship was positive or negative.~~**

~~5~~4 Discussion

4.1 Monitoring setup and associated uncertainties

Regarding chemographs and calculation of R_C, uncertainties arose from timing and frequency of sampling and analytical error, and additionally from discharge measurement when calculating masses and R_M. Analytical methods used in this study usually produced very consistent results so that variability in concentrations of parent compounds in duplicate samples was low (sd < 10 %). However, in individual samples collected at G1 analytical variability was elevated for met-ESA and met-OA (Figure 3), reducing confidence in concentrations and the derived measures R_C and R_M of TPs in the affected chemographs (E2, E5, E7, E8, E9). Uncertainty related to timing and frequency of sampling can hardly be quantified but certainly depends on how well the sampling intervals captured variability in concentrations during flood events and how well the time lag between upstream and downstream sampling matched the residence time of solutes in the wetland. Lefrancq et al. (2017) assessed the effect of sampling frequency in pesticide monitoring data collected during runoff from a single vineyard and found that acute toxicity of pesticide flushes was underestimated up to 4-times when calculated from event means and up to 30-times when calculated from random samples. Although these data were collected on the plot scale and we assume that variability in our catchment is lower due to longer flow paths and mixing processes on the catchment scale, uncertainty of the chemographs in our study could have been reduced by increasing sampling frequency. Regarding the timing of upstream and downstream sampling, there is evidence that water residence time in the wetland was in fact shorter than one hour. The observation that for quickly responding compounds, such as boscalid, concentration in the first sample at G2 was often elevated compared to the first sample at G1 indicates that the contaminant flush hat already reached G2 when sampling started. This did not influence determination of $C_{out,max}$ and R_C in the outlet of the wetland, as concentrations were still rising from the first to the second sample (Figure 3). However, effects on M_{out} were higher, since a relevant fraction of contaminant mass leaving the wetland was not registered and thereby caused overestimation of R_M (Figure S1). Another source of uncertainty exclusively affecting contaminant mass and not concentrations was the use of different gauging systems at G1 and G2. Different shapes of the measurement cross-section (triangular at G1 and rectangular at G2) caused G2 to be less precise and water imbalances on the event scale, particularly when flow was low. Summarizing the setup constraints above, we have high confidence that the experimental setup produced realistic chemograph shapes and captured peak concentration reasonably well, but are less confident regarding contaminant loads.

4.2 Mobilization of contaminants and formation of distinct chemographs

Peak concentrations of mobilized contaminant flushes were different depending on the compound. This may be due to the application of different amounts and due to temporal patterns of application. The fact that maximum concentrations of metazachlor and flufenacet in specific events exceeded concentrations during most other events by a factor of more than 100, suggested application of these compounds shortly before the onset of runoff. Despite highly variable application patterns, our cluster analysis resulted in four groups with similar chemograph shape. Many factors have been shown in literature to influence the mobilization of pesticides in catchments, including catchment properties, event properties and physiochemical compound properties. As catchment properties we here consider factors associated with runoff generation such as catchment geometry, terrain slopes, and in particular the delineation of areas where different compounds were applied. The interplay of these factors defines hydrological activity and connectivity (i.e. by shortcuts like roads and drainage pipes) of critical source areas for different compounds (Doppler et al., 2012; Gomides Freitas et al., 2008). Event properties include intensity and dynamics of rainfall (Imfeld et al., 2020) and subsequent runoff (Doppler et al., 2014). Relevant physiochemical compound properties are e.g. mobility and degradability (Gassmann et al., 2015).

These properties are reflected to varying degrees in the results of the cluster analysis. Cluster A was characterized by a quick response and a concentration plateau towards the end of sampling and was mainly composed of TPs. The fact that concentration maxima in cluster A were delayed compared to fungicides (cluster B), although their parent compounds were applied closer to the stream in the flat valley bottoms, suggests that they were transported with a slower flow components. Due to flatter terrain, surface runoff played a less important role and the main transport pathway was subsurface flow. Where fields were undrained, however, transit time of water from the infiltration point to the stream would likely exceed the temporal scale of event sampling. Most of the water reaching the stream from the fields in the valley during discharge events would therefore be pre-event water, enriched in TPs formed in the soil, corresponding to the formation site of TPs of the chloracetamide herbicides to which metazachlor belongs (Mersie et al., 2004). Seepage of pre-event TP-rich water thus explains the immediate response of chemographs in cluster A. The quick response was often followed by a local concentration minimum between samples 2 and 5, i.e. between 30 min and 6 h after sampling was initialized. Coincidence of this minimum with concentration peaks of fungicides might suggest dilution of TP concentration by mixing with event water carrying high loads of fungicides but less TPs of metazachlor.

Cluster B represented differences between fungicides and the remaining compounds. Considering land use distribution in the studied catchment, it is unclear whether this partition reflects different compound properties or catchment properties or both. The fact that concentration in cluster B quickly increased with discharge (within 30 minutes) is in line with fast transport from the vineyard terraces to the stream via roads and drainage pipes as described by Gassmann et al. (2012) for suspended solids in the studied catchment. Along such preferential pathways, compound properties, such as sorption affinity, may be less important (Gomides Freitas et al., 2008) compared to e.g. percolation through the soil with intense contact to sorption sites in the soil matrix. Moreover, fungicides are applied by sprayers into the foliage and can drift to e.g. paved surfaces from which they can be quickly mobilized by subsequent rainfall (Lefrancq et al., 2013). We therefore hypothesize that cluster B was mainly produced by surface flushing and fast transport pathways of fungicides. This explained the quick rise and subsequent decline in concentrations (concurrent with plateaus produced by slower flow components in cluster A).

Cluster C was composed of chemographs of all compounds but mainly from events E2 and E4 indicating event dependence. Two aspects were found to support this idea. First, there was a secondary discharge peak in event 2 that did not contribute much in terms of contaminant concentration but rather caused dilution and produced particularly flat chemograph tails. Second, peaks of herbicides and TPs were less delayed compared to fungicides. This may be the result of recent herbicide application and active surface runoff in the flat valleys. Timing of pesticide application was identified as the main export driver of currently used pesticides by Imfeld et al. (2020) who performed a cluster analysis on rainfall data from a headwater vineyard catchment. Based on the magnitude of discharge and amount of mobilized contaminants (concentration of metazachlor ≈ 10 µg L^{-1}), both explanations seem plausible in event E2. Event E4, however, did not show particularly high herbicide concentration nor a secondary discharge peak. Although it is obvious that chemograph shapes in cluster C differed from the other clusters, unfortunately, the responsible factors remain unclear.

Cluster D included chemographs of both herbicides and TPs and presented a clear peak that was often defined by a single sample 2 h after the beginning of the event. In contrast to cluster A, cluster D was characterized by a single sharp discharge peak (except in event E7 where a second peak occurred shortly after the first) and mainly included chemographs during periods of low flow. Our interpretation is that cluster D represented flow events in which no dilution of herbicide and TP fluxes by fungicide fluxes or secondary discharge peaks occurred. Low pre-event discharge in cluster D compared to cluster A may indicate low water levels which may have caused a slower response as no enriched pre-event water was released from the soils in the valleys.

The unclear interpretation of cluster C suggest that we missed important factors for the formation of chemographs. In fact, variables like spatial distribution of rainfall or pesticide application rates and timing (Imfeld et al., 2020) and possibly other factors likely influenced chemograph shapes. Knowing all these variables would not change the results produced by the clustering algorithm but rather increase our ability to interpret them. Nevertheless, the cluster analysis helped to explore how the catchment and processes therein influenced concentration signals of mobilized contaminants. Particularly, the analysis helped to understand under what conditions and for which pollutant sharp-peaked chemographs, associated with high acute toxicity, can be expected. We therefore see a high potential of this type of analysis for the identification of influential factors for contaminant mobilization in other catchments, although these factors may not be universal but catchment-dependent.

4.3 Mitigation efficiency and chemograph shape

4.3.1 Peak concentration reduction

We hypothesized that peak concentration reduction in the VTS will be highest for chemographs with the sharpest peaks, i.e. for the chemographs that were most sensitive to dispersion. And indeed we found a systematic relationship between R_C and both i_{DS} and chemograph clusters. Although the relationship of clusters and R_C largely reflected the relationship between R_C and i_{DS}, it is surprising that R_C was clearly highest in cluster D and not in cluster C which presented better defined peaks and slightly higher i_{DS} per cluster (Figure 4f). Critical inspection of input chemographs shows that in several chemographs of TPs (met-ESA and met-OA in event E4 and met-OA in event E8) elevated concentrations in the last samples exhibited high analytical errors and did not appear in the outlet chemograph. These dubious samples caused low i_{DS} but substantial R_C and thus contributed to variability in i_{DS} despite high values of R_C in cluster D. We therefore do not consider the deviation from the expected cluster ordering contradictory but to result from increased uncertainty in cluster D as mentioned earlier. In contrast, the hypothesized relationship between RC and chemograph shape was demonstrated for both i_{DS} and chemograph clusters, the latter of which also integrates shape aspects that go beyond i_{DS}, e.g. timing of peaks. Overall, the values of R_C found in our study compare with field data from vegetated buffers (Bundschuh et al., 2016; Stehle et al., 2011) and are in the range of those found in vegetated stream mesocosms by Elsaesser et al. (2011) and Stang et al. (2014) who both attributed most of the observed peak reduction to dispersion.

In addition, we found relationships between and R_C and discharge dynamics, i.e. Q_{mea} and ratio of Q_{max} to Q_{mean}. The influence of discharge on R_C may be two-fold. First, increasing flow reduced residence time and hydraulic efficiency, i.e. short circuiting reduced the potential for dispersion and interaction with wetland sediments or plants. Second, the fact that chemographs of events with high Q_{max} to Q_{mean} ratios were attributed to cluster D suggests that discharge dynamics influenced the shape of the chemograph at the wetland inlet. This means, the influence of discharge may also be indirect by promoting the formation of sharp-peaked chemographs with high potential for peak reduction.

In contrast to other studies, we did not find clear relationships of R_C to and physiochemical properties of compounds such as sorption affinity (Stehle et al., 2011; Vymazal and Březinová, 2015) or solubility (Bundschuh et al., 2016). The absence of such relationships may partially be due to the low number of different target compounds in our study (n=6). However, given the short time lag between sampling at the inlet and outlet of the wetland (Δt = 1h), it seems logical that no relevant sorption or degradation occurred within this period. For comparison, in batch experiments by Gaullier et al. (2018) adsorption equilibrium for boscalid (compound with second highest Kf_{OC} in our study) was only reached after 24 h. Despite the relatively narrowly confined R_C values of the parent compounds, we do not consider physiochemical compound properties as major drivers of R_C in our VTS.

4.3.2 Contaminant mass removal

For R_M we found a different pattern among the chemograph clusters than for R_C. R_M was apparently higher in clusters A and D than in clusters B and C. However, the clusters indicating substantial mass removal were those with increased uncertainty regarding compound mass. Cluster A often showed relevant mass flux at the end of sampling (and presumably beyond) which we did not account for. Cluster D contained dubious data points of TPs and poorly defined peaks outside the periods of high sampling frequency. In addition, due to overestimation of solute travel time in the wetland in the monitoring setup, the rising limp of the mass flux signal at G2 was often not adequately captured by the sampling scheme, causing underestimation of downstream event mass and overestimation of mass loss. In absence of any clear relationship with compound properties, discharge dynamics or chemograph shape, this suggests that the assessment of contaminant masses was subject to systematical errors and that the apparent mass loss found in our study should therefore not be over-interpreted.

In earlier studies, Lange et al. (2011) and Schuetz et al. (2012) observed a 15-30 % mass loss of the fluorescent tracer sulforhodamine-B in the wetland subsection of the current VTS. These results indicate a general potential for sorption of organic compounds in this system, but represent an earlier succession state of the wetland and stationary flow conditions with much longer residence times. Also in the current VTS kinetic sorption of contaminants may have occurred but sorption equilibrium was certainly not reached (Gaullier et al., 2018). Thus the effect of sorption did not reach its full potential. In fact, other studies reported limited mass removal in wetlands with comparable residence times. Ramos et al. (2019) did not find relevant R_M in two surface flow wetlands with residence times between 45 min and 6 h in England. In contrast, Passeport et al. (2013) found R_M between 45 % and 96 % in a constructed wetland with a residence time of 66.5 h. However, their contaminant mass loss coincided with loss of water (45 %). Mesocosm experiments by Elsaesser et al. (2011) and Stang et al. (2014) showed strong concentration reduction but only very limited and temporary mass removal at residence times of a few hours. In summary, these findings suggest that the potential for mass removal in wetland systems like the one studied here is rather limited. However, wetlands have been shown to reduce contaminant mass, when residence times are sufficiently long (Gregoire et al., 2009) or when operated in batch mode (Tournebize et al., 2017; Moore et al., 2000; Maillard et al., 2016).

~~6 Discussion~~

~~Concentrations detectable at the catchments may depend on many factors including stream size~~ (Lorenz et al., 2017) ~~and land-use which complicates direct comparison. Peak concentrations of fungicides detected in this study, however, were in the range~~

~~of those detected by~~ Bundschuh et al. (2016) ~~in different streams draining viticultural catchments in Southwest Germany. The fact that highest peak concentrations were detected for the fungicides reflects both the land-use distribution in the catchment (higher percentage of vineyards compared to cropland) and the different application practices (higher application frequencies of fungicides). The increase in PC to TP concentration ratio from stationary to transient flow conditions indicates that TPs are exported via a permanently active pathway such as groundwater. In contrast, PCs are rather transported via periodically active pathways that respond quickly to rainfall. This interpretation is in line with the patterns identified in section 2.6 and findings of~~ Gassmann et al. (2013)~~.~~

4.4 Conclusions

In agreement with other studies this investigation shows that VTSs with short water residence times of up to several hours may cause substantial reduction of peak concentrations of contaminants mobilized during discharge events. This implies an efficient reduction of acute toxicity for receiving aquatic ecosystems. In the present VTS the reduction of concentration peaks was mainly controlled by dispersion and was more pronounced for sharp-peaked than for flat input chemographs. In contrast, contaminant mass loss was rather limited, mainly due to the fact that short residence times did not allow for considerable sorption or transformation. Clustering of chemographs revealed that chemograph shapes were associated with source areas, input pathways and discharge dynamics. This highlighted the role of chemographs as links between processes in catchments and in receiving aquatic systems. The presented cluster analysis helped to understand why and for which pollutant sharp-peaked chemographs could be expected. Such sharp-peaked chemographs produce high acute toxicity in aquatic ecosystems but at the same time can efficiently be mitigated in VTSs. While the factors controlling chemograph shape may be different in different catchments, the effect dispersion exerts on these signals is universal.

~~As the idea of dispersion sensitivity is inherent in the concept of advective-dispersive transport, we consider our findings regarding the role of chemograph shape in principle transferable to other systems. However, the factors contributing to chemograph shapes may be different in different catchments. The implication of our finding towards the design of treatment systems are that residence should be maximized if degradation of pesticides is desired. However, residence time should also clearly exceed the duration of discharge events, if peak reduction is desired, so that dispersion and mixing processes can happen in the wetland. chemograph shapes of mobilized contaminants can be attributed to contaminant groups and their source areas. Both comparison of absolute reduction rates and regression analysis suggest that the shape of the input signal may play an important role for peak concentration reduction in VTSs with short HRT. We therefore recommend that this factor should be considered for future assessment of VTS functionality and ecotoxicicty based risk assessment. Our findings imply that peak concentration reduction in VTSs may generally be more efficient for compounds As dispersion sensitivity is inherent in the concept of advective-dispersive transport, we consider these findings to be transferable to other solutes mobilized in catchments during storm events.whose chemographs are characterized by sharp peaks and low backgrounds rather than for those ubiquitous in a catchment.~~

Data availability

Contaminant data used in this study is available as supplementary material.

Author contributions

JL planned the monitoring concept. JG performed the data analysis and prepared ~~the~~ most of the manuscript in cooperation with JL. ~~with contributions from all co-authors.~~ OO and KK facilitated the sample analysis and OO wrote the section on analytical methodology. ~~JG performed the data analysis and prepared the manuscript with contributions from all co-authors.~~

Competing interests

The authors declare that they have no conflict of interest.

Financial support

This research has been supported by the Federal Ministry of Education and Research (BMBF) (grantno. 02WRM1366B) and the Water Network Baden-Württemberg funded by the Ministry of Science, Research and Art of the State of Baden-Württemberg. The article processing charge was funded by the Baden-Wuerttemberg Ministry of Science, Research and Art and the University of Freiburg in the funding program Open Access Publishing.

References

Aubert, A. H. and Breuer, L.: New seasonal shift in in-stream diurnal nitrate cycles identified by mining high-frequency data, PLoS One, 11, e0153138, doi:10.1371/journal.pone.0153138, 2016.

Beketov, M. A., Kefford, B. J., Schäfer, R. B., and Liess, M.: Pesticides reduce regional biodiversity of stream invertebrates, Proc. Natl. Acad. Sci. USA, 110, 11039–11043, doi:10.1073/pnas.1305618110, 2013.

Bos, M. G. (Ed.): Discharge measurement structures, International Institute for Land Reclamation and Improvement, Wageningen, The Netherlands, 1989.

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.

Doppler, T., Camenzuli, L., Hirzel, G., Krauss, M., Lück, A., and Stamm, C.: Spatial variability of herbicide mobilisation and transport at catchment scale: insights from a field experiment, Hydrol. Earth Syst. Sci., 16, 1947–1967, doi:10.5194/hess-16-1947-2012, 2012.

Doppler, T., Lück, A., Camenzuli, L., Krauss, M., and Stamm, C.: Critical source areas for herbicides can change location depending on rain events, Agr. Ecosyst. Environ., 192, 85–94, doi:10.1016/j.agee.2014.04.003, 2014.

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.

Fenner, K., Canonica, S., Wackett, L. P., and Elsner, M.: Evaluating pesticide degradation in the environment: Blind spots and emerging opportunities, Science, 341, 752–758, 2013.

Fernández, D., Voss, K., Bundschuh, M., Zubrod, J. P., and Schäfer, R. B.: Effects of fungicides on decomposer communities and litter decomposition in vineyard streams, Sci. Total Environ., 533, 40–48, doi:10.1016/j.scitotenv.2015.06.090, 2015.

Fischer, H. B., List, J. E., Koh, C. R., Imberger, J., and Brooks, N. H.: Mixing in inland and coastal waters, Academic Press, San Diego, California, 302 pp., 1979.

Gassmann, M., Lange, J., and Schuetz, T.: Erosion modelling designed for water quality simulation, Ecohydrol., 5, 269–278, doi:10.1002/eco.207, 2012.

Gassmann, M., Olsson, O., Stamm, C., Weiler, M., and Kümmerer, K.: Physico-chemical characteristics affect the spatial distribution of pesticide and transformation product loss to an agricultural brook, Sci. Total Environ., 532, 733–743, doi:10.1016/j.scitotenv.2015.06.068, 2015.

Gassmann, M., Stamm, C., Olsson, O., Lange, J., Kümmerer, K., and Weiler, M.: Model-based estimation of pesticides and transformation products and their export pathways in a headwater catchment, Hydrol. Earth Syst. Sci., 17, 5213–5228, doi:10.5194/hess-17-5213-2013, 2013.

Gaullier, C., Dousset, S., Billet, D., and Baran, N.: Is pesticide sorption by constructed wetland sediments governed by water level and water dynamics?, Environ. Sci. Pollut. Res. Int., 25, 14324–14335, doi:10.1007/s11356-017-9123-1, 2018.

Gomides Freitas, L., Singre, H., Müller, S. R., Schwarzenbach, R. P., and Stamm, C.: Source area effects on herbicide losses to surface waters—A case study in the Swiss Plateau, Agric. Ecosyst. Environ., 128, 177–184, doi:10.1016/j.agee.2008.06.014, 2008.

Gregoire, C., Elsaesser, D., Huguenot, D., Lange, J., Lebeau, T., Merli, A., Mose, R., Passeport, E., Payraudeau, S., Schütz, T., Schulz, R., Tapia-Padilla, G., Tournebize, J., Trevisan, M., and Wanko, A.: Mitigation of agricultural nonpoint-source pesticide pollution in artificial wetland ecosystems, Environ. Chem. Lett., 7, 205–231, doi:10.1007/s10311-008-0167-9, 2009.

Grömping, U.: Relative importance for linear regression in R: The package relaimpo, Journal of Statistical Software, 17, 2006.

Hair, J. F.: Multivariate data analysis, 7. ed., Pearson Prentice Hall, Upper Saddle River, NJ, 785 pp., 2010.

Han, J., Kamber, M., and Pei, J.: Data mining: Concepts and techniques, 3. ed., The Morgan Kaufmann series in data management systems, Elsevier/Morgan Kaufmann, Amsterdam, 703 pp., 2012.

Hartigan, J. A. and Wong, M. A.: Algorithm AS 136: A k-means clustering algorithm, Applied Statistics, 28, 100, doi:10.2307/2346830, 1979.

Hensen, B., Olsson, O., and Kümmerer, K.: A strategy for an initial assessment of the ecotoxicological effects of transformation products of pesticides in aquatic systems following a tiered approach, Environ. Int., 137, 105533, doi:10.1016/j.envint.2020.105533, 2020.

Imfeld, G., Meite, F., Wiegert, C., Guyot, B., Masbou, J., and Payraudeau, S.: Do rainfall characteristics affect the export of copper, zinc and synthetic pesticides in surface runoff from headwater catchments?, Sci. Total Environ., 741, 140437, doi:10.1016/j.scitotenv.2020.140437, 2020.

Lange, J., Schuetz, T., Gregoire, C., Elsässer, D., Schulz, R., Passeport, E., and Tournebize, J.: Multi-tracer experiments to characterise contaminant mitigation capacities for different types of artificial wetlands, Int. J. Environ. An. Ch., 91, 768–785, doi:10.1080/03067319.2010.525635, 2011.

Lefrancq, M., Imfeld, G., Payraudeau, S., and Millet, M.: Kresoxim methyl deposition, drift and runoff in a vineyard catchment, Sci. Total Environ., 442, 503–508, doi:10.1016/j.scitotenv.2012.09.082, 2013.

Lefrancq, M., Jadas-Hécart, A., La Jeunesse, I., Landry, D., and Payraudeau, S.: High frequency monitoring of pesticides in runoff water to improve understanding of their transport and environmental impacts, Sci. Total Environ., 587-588, 75–86, doi:10.1016/j.scitotenv.2017.02.022, 2017.

Lewis, K. A., Tzilivakis, J., Warner, D. J., and Green, A.: An international database for pesticide risk assessments and management, Hum. Ecol. Risk Assess., 22, 1050–1064, doi:10.1080/10807039.2015.1133242, 2016.

Lindemann, R. H., Merenda, P. F., and Gold, R. Z.: Introduction to bivariate and multivariate analysis, 3rd ed., Scott Foresman, Glenview, 444 pp., 1980.

Lorenz, S., Rasmussen, J. J., Süß, A., Kalettka, T., Golla, B., Horney, P., Stähler, M., Hommel, B., and Schäfer, R. B.: Specifics and challenges of assessing exposure and effects of pesticides in small water bodies, Hydrobiologia, 793, 213–224, doi:10.1007/s10750-016-2973-6, 2017.

Lumley, T.: leaps: Regression Subset SelectionRegression: R package, 2017.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K.: cluster: Cluster Analysis Basics and Extensions. R package, 2019.

Maillard, E. and Imfeld, G.: Pesticide mass budget in a stormwater wetland, Environ. Sci. Technol., 48, 8603–8611, doi:10.1021/es500586x, 2014.

Maillard, E., Lange, J., Schreiber, S., Dollinger, J., Herbstritt, B., Millet, M., and Imfeld, G.: Dissipation of hydrological tracers and the herbicide S-metolachlor in batch and continuous-flow wetlands, Chemosphere, 144, 2489–2496, doi:10.1016/j.chemosphere.2015.11.027, 2016.

Mersie, W., McNamee, C., Seybold, C., Wu, J., and Tierney, D.: Degradation of metolachlor in bare and vegetated soils and in simulated water-sediment systems, Environ. Toxicol. Chem., 23, 2627–2632, doi:10.1897/04-60, 2004.

Moore, M. T., Rodgers Jr., J. H., Cooper, C. M., and Smith Jr., S.: Constructed wetlands for mitigation of atrazine-associated agricultural runoff, Environ. Pollut., 110, 393–399, doi:10.1016/s0269-7491(00)00034-8, 2000.

Oliver, D. P., Kookana, R. S., Anderson, J. S., Cox, J. W., Waller, N., and Smith, L. H.: Off-site transport of pesticides in dissolved and particulate forms from two land uses in the Mt. Lofty Ranges, South Australia, Agr. Water Manage., 106, 78–85, doi:10.1016/j.agwat.2011.11.001, 2012.

Olsson, O., Khodorkovsky, M., Gassmann, M., Friedler, E., Schneider, M., and Dubowski, Y.: Fate of Pesticides and Their Transformation Products: First Flush Effects in a Semi-Arid Catchment, Clean (Weinh), 41, 134–142, doi:10.1002/clen.201100545, 2013.

Passeport, E., Tournebize, J., Chaumont, C., Guenne, A., and Coquet, Y.: Pesticide contamination interception strategy and removal efficiency in forest buffer and artificial wetland in a tile-drained agricultural watershed, Chemosphere, 91, 1289–1296, doi:10.1016/j.chemosphere.2013.02.053, 2013.

Peter, G.: Überfälle und Wehre: Grundlagen und Berechnungsbeispiele, Vieweg Verlag, Wiesbaden, 320 pp., 2005.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2019.

Ramos, A., Whelan, M. J., Guymer, I., Villa, R., and Jefferson, B.: On the potential of on-line free-surface constructed wetlands for attenuating pesticide losses from agricultural land to surface waters, Environ. Chem., 16, 563–576, doi:10.1071/EN19026, 2019.

Reichenberger, S., Bach, M., Skitschak, A., and Frede, H.-G.: Mitigation strategies to reduce pesticide inputs into ground- and surface water and their effectiveness; a review, Sci. Total Environ., 384, 1–35, doi:10.1016/j.scitotenv.2007.04.046, 2007.

Schuetz, T., Gascuel-Odoux, C., Durand, P., and Weiler, M.: Nitrate sinks and sources as controls of spatio-temporal water quality dynamics in an agricultural headwater catchment, Hydrol. Earth Syst. Sci., 20, 843–857, doi:10.5194/hess-20-843-2016, 2016.

Schuetz, T., Weiler, M., and Lange, J.: Multitracer assessment of wetland succession: Effects on conservative and nonconservative transport processes, Water Resour. Res., 48, 97, doi:10.1029/2011WR011292, 2012.

Stang, C., Wieczorek, M. V., Noss, C., Lorke, A., Scherr, F., Goerlitz, G., and Schulz, R.: Role of submerged vegetation in the retention processes of three plant protection products in flow-through stream mesocosms, Chemosphere, 107, 13–22, doi:10.1016/j.chemosphere.2014.02.055, 2014.

Stehle, S., Elsaesser, D., Gregoire, C., Imfeld, G., Niehaus, E., Passeport, E., Payraudeau, S., Schäfer, R. B., Tournebize, J., and Schulz, R.: Pesticide risk mitigation by vegetated treatment systems: a meta-analysis, J. Environ. Qual., 40, 1068–1080, doi:10.2134/jeq2010.0510, 2011.

Taghavi, L., Merlina, G., and Probst, J.-L.: The role of storm flows in concentration of pesticides associated with particulate and dissolved fractions as a threat to aquatic ecosystems - Case study: the agricultural watershed of Save river (Southwest of France), Knowl. Managt. Aquatic Ecosyst., 6, doi:10.1051/kmae/2011002, 2011.

Tournebize, J., Chaumont, C., and Mander, Ü.: Implications for constructed wetlands to mitigate nitrate and pesticide pollution in agricultural drained watersheds, Ecol. Eng., 103, 415–425, doi:10.1016/j.ecoleng.2016.02.014, 2017.

Vymazal, J. and Březinová, T.: The use of constructed wetlands for removal of pesticides from agricultural runoff and drainage: a review, Environ. Int., 75, 11–20, doi:10.1016/j.envint.2014.10.026, 2015.

Zubrod, J. P., Bundschuh, M., Arts, G., Brühl, C. A., Imfeld, G., Knäbel, A., Payraudeau, S., Rasmussen, J. J., Rohr, J., Scharmüller, A., Smalling, K., Stehle, S., Schulz, R., and Schäfer, R. B.: Fungicides: An Overlooked Pesticide Class?, Environ. Sci. Technol., 53, 3347–3365, doi:10.1021/acs.est.8b04392, 2019.