

## Response to Reviews

To all reviewers

Thanks for the detailed and constructive comments. Below we added our responses (in red) to each single point.

On behalf of all co-authors.

Christian Stamm

Anonymous Referee #1

General comments

*Developping and testing the limits of a parsimonious model of micropollutants transport from catchment to river at various scales is relevant both for stakeholders and for the scientific community to address the degree of simplification required/able to capture the micropollutants patterns. The key concept of this approach (link the load of micropollutants to the discharge and/or the rain) is not new but the approach to validate at small scale with European databases the load then upscaling the approach at larger scale reaching the main part of the Rhine catchment is new. The spatial preprocessing of existing European data to improve the information for the subcatchment can also be underlined. The state of the art on the different components covered by the work is well presented with relevant references. The model development, main hypotheses and calibration/validation/transposition steps are clearly presented.*

**Thanks for the positive feedback.**

*However, a scheme summarizing spatial and temporal discretization with associated processes across scale (calibration then validation catchments and full Rhine scale) is clearly missing in the main text. A very simple scheme (focused on subcatchment delineation) is presented in SI but can't play this summary goal.*

**We agree that such a scheme would be useful and will include an improved version of the current figure in the SI into the main text.**

*The results are clearer for herbicides than for biocides. The reader discovered different hypotheses that reduce progressively the extent of the biocides loads ant transport modelling at larger scale. I wonder if a focus on herbicides only, should not be better and stronger.*

**We agree that the data basis for modelling biocides is substantially poorer than for herbicides. However, we consider biocides as relevant compounds and find it thus important to report the modelling challenges and thereby illustrate the limitations when simulating such compounds.**

*I detailed below specific comments and corrections*

*P1 L30: check homogeneity (S-metolachlor and metolachlor are used in the text)*

Thanks for the careful reading. We will check for consistency, use consistent terminology and make it explicit if one has to switch to the other expression because data do not allow the specific use of S-metolachlor.

*P2 L28: Missing dot at the end of a sentence "2013) One"*

Will be corrected.

*P3 L13-14: original aspect of this work +add that it's a daily or hourly time step*

We will add the temporal resolution.

*P3 L23: all the basin? just after it's mentioned that only the basin upstream of the station Emmerich am Rhein is covered by this study, precise*

We will mention here that the basin upstream of Emmerich is covered.

*P4 L6: general comment: An overall scheme of the model should be relevant to improve understanding of the spatial links between objects, considered processes and links between AWaQa and AQUASIM (what's happened in the sub-catchment scale and then in the river, degradation, ... ) The appendix A2 is one aspect of the discussion but it not covers the processes.*

As mentioned above, we agree and will add a corresponding scheme.

*P5 L8 and L9: in the equations 2 and 3 Kdeg has to be used instead of Kd to be homogeneous with the rest of the paper*

Thanks for noticing. This will be corrected.

*P5 L11: not clear at this stage how the available fraction is link to rainfall*

This fraction is not linked to rainfall but describes the effect of immediate sorption in contrast to the slower exchange processes described by Eq. [3]. We try to rephrase such to avoid misunderstandings.

Suggested sentence: "... available for transport, such that it can be directly mobilised when it rains."

*P5 L21: interesting but why 1/14? Expert panel, reasonable fractionation? Sensitivity of the model to this fraction? Does it means that 14 days are required to consider 100% of application? Can this hypothesis impact peak modelling due to dilution of input signal?*

Yes, the total application mass is distributed over 14 days. This reflects the fact that pesticides are not applied on all fields in a subcatchment on the same day. With applications taking place only during dry days, the total application period lasts 3 – 4 weeks which is reasonable based on our field experience.

These assumptions have of course an influence on simulated peak concentrations because they only reflect an approximation to the actual temporal application pattern, which is unfortunately not known.

*P6 L23: no rainfall dependency?*

Yes, the load is proportional to rainfall - the last term (P(t)) represents daily precipitation.

P6 L34-35: not clear for me, does it mean that any transformation is considered in the river routing model?

Yes, the load aggregation models neglects transformation processes while the routing procedure with Aquasim in principle could.

P7 L8: should be interesting to mention in the abstract and in the introduction...

We will mention the coupling with Aquasim in both parts.

P7 L11-12: is it possible to mention an unpublished paper in HESS?

We will skip this reference and the one on page 16

P7 L17: I think it is the table S2 (in the appendix A4) and not the table S4.

Thanks for noticing. This will be corrected.

P7 L20: not easy to understand even if it's describe below... a scheme should be relevant to improve understanding.

We will add a schematic representation for better illustration.

P7 L25: what % considered considering strahler order less than five?

We will add the respective information.

P7 27: Explain briefly the area ratio method to help the reader at this stage

The area ratio method simply assumes that discharge scales proportional to catchment area (see e.g. Hirsch 1979). We will add this explanation to the sentence: "..., which assumes that discharge scales proportional to catchment area." .

P7 31: Same comment, explain briefly the map-correlation method to help the reader at this stage.

We will add a sentence for more explicit explanation: "Selection of the reference stations is based on the map-correlation method from Archfield and Vogel. (2010). This geostatistical method calculates the correlation between discharge time series at observed stream gauges and estimates the station with the most correlated discharge at the ungauged catchment based."

P8 L16-18: not clear for me!  $-6.5$  per km (in z?)

We will rephrase ("From these altitude deviations, temperature values were corrected based on a temperature decrease by  $-0.0065^{\circ}\text{C}/\text{m}$  altitude increase".)

P8 L21: Expected evolution if 2000 is the reference? Impact of rotation of crops (spatial difference between years)?

This is a very important point and there are two aspects to this question. The first relates to the long-term development of cropping patterns. It is unfortunately true that the available data base on spatial cropping patterns does not reflect the most recent situation. In order to check whether there have been major shifts in agricultural land use, we inspected the (spatially lumped) temporal

evolution of cropping areas for the different countries and the relevant crops (maize, wheat, sugar beet) based on the FAO statistics (<http://www.fao.org/faostat/en/#data>; accessed 26 March 2018).

These aggregated data reveal only slight changes in the planting of these major crops over the last 20 years. This supports our assumption that the spatial patterns have not changed much and that our land use data adequately reflect land use for our study period. We will add an explanatory sentence in the main text and add the figure on the temporal trends since 1961 in the SI (see Fig. 1 below).

In this context, we realised that we were not explicit about the crops on which the different herbicides are used. We will add this information in Appendix A4, Table S2.

The second aspect to this comment relates to the crop rotation, which shifts crops between fields for different years. However, these spatial shifts occur at a very local scale and average out at scales relevant for this study. An exception may be the very small catchments (1 – 2 km<sup>2</sup>) used for model calibration. In these cases however, we rely on specific and detailed land use data for the specific year.

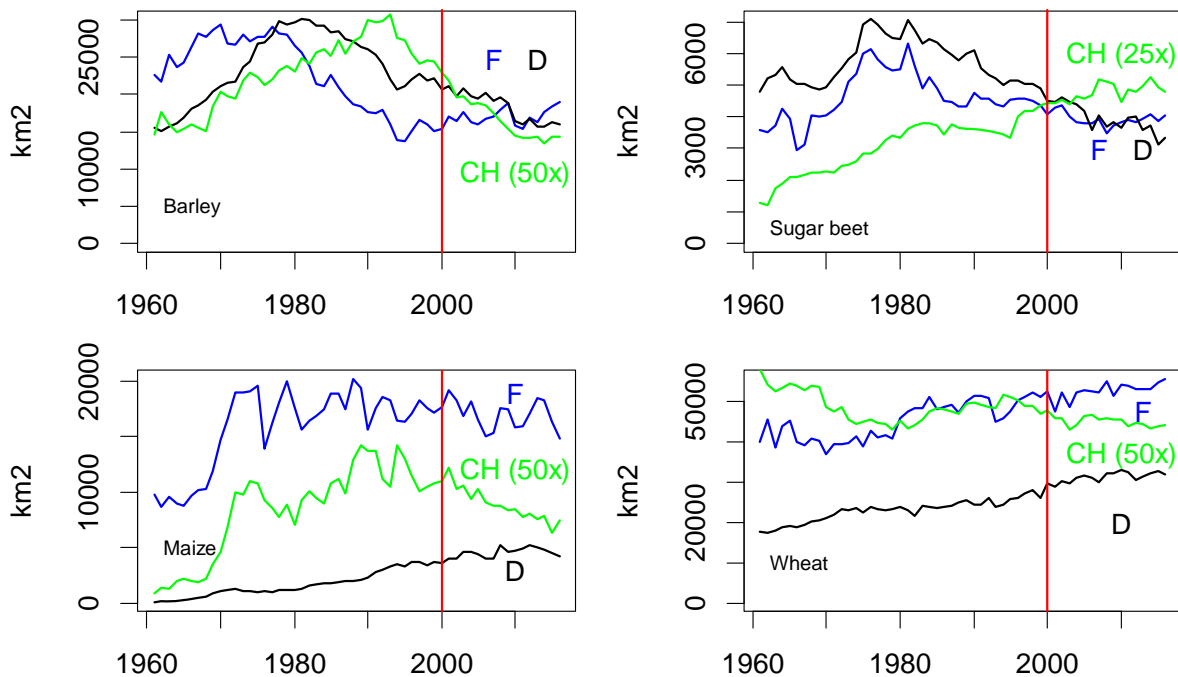


Fig. 1: Temporal evolution of cropping area for four major crops relevant for the modelled herbicide use. The red line indicates the reference year for the spatial cropping patterns. Data source: <http://www.fao.org/faostat/en/#data>.

P9 L17: Diuron was also used in vineyard still 2008 in France. Is it the same in Germany and Switzerland? Expected impact of missing the agricultural uses in this analysis?

In Switzerland, diuron is indeed still registered and used in vineyards. However, the areal coverage in our study area is of minor importance and therefore neglected.

In Germany, diuron is not registered as agricultural herbicide.

In France, diuron was banned in 2007 with the possibility to use existing stocks by the end of 2008.

In summary, we could safely neglect diuron applications in agriculture for our study area and study period.

P9 L24-29: not clear for me: range of available data? 2008-2012 so 5 years for all the study site excepted for Lorraine? Right? You have to mention the Appendix A5 (figure S3 and S4) to visualize the spatial variability. The reader discovers in the caption of the figure S4 (appendix A5) that the Diuron pattern was just a copy of the carbendazim map. Is it correct? I do not find any discussion on that in the text.

We will refer in the main text to figures S3 and S4 for the spatial patterns.

We will also mention that the spatial diuron and carbendazim patterns are identical because of having the same source areas and uniform application rates across space because there is no data that allows for spatial differentiation.

P9 L34: does any other sources can (even partially) validate this hypothesis?

Unfortunately, we are not aware of any such data source.

P10 L22: "application season" Perhaps explain why to take into account the fact that during application period difference can occur between real applications (unknown) and modelled application (splitted with 1/14 depending of weather windows).

We will add a further explanation.

P10 L22: "error-scaling function" try to better introduce this function, why and how, it's not very clear for the reader.

We will add a further explanation: "The error scaling function makes the standard deviation of herbicide errors proportional to the remaining field stock to reflect that errors are larger in the application period than afterwards, when the compound is present in negligible amounts." This was necessary because the applied Box-Cox transformation alone could not achieve the normality of residuals.

P10 L30: the text and the equation 13 are not supported by any reference

Power transformation of input data and model results is common practice to ensure the normality of residuals. The Box-Cox transformation (Box and Cox 1964) is the most common form of transformation applied for this purpose. We will refer to the iWaQa study (Honti et al. 2017) as a precursor. Equation (13) is a plain normal likelihood function accounting for transformation (Box and Cox 1964).

P11 L10: studies are mentioned to assess the prior distribution of  $\varepsilon$ , which one? Wittmer et al (2010) mentioned in L26?

We will rephrase the sentence to "... from these studies.." such that is clear that we refer to the studies mentioned on p. 11, L. 8 – 9 and listed in Appendix A8, Tab. S4).

P11 L28: I suggest to mention the table S4 and S5 (Appendix A8) after "of the priors for  $\varepsilon$  and  $\beta$ "

We will add this information.

P12 L28: “larger rivers” (Rhine, Aare;” I suggest to mention all main tributaries or use “such as”

We will modify the text accordingly.

P12 L31: The fact that biocide was finally not modelled at the Rhine scale (due to lack of biocides export coefficient in France and Germany) should be precise clearly in the abstract and introduction by differencing the two scales (calibration/validation in the Switzerland scale of herbicides and biocides, and only extrapolation of herbicides at the Rhine scale).

We will make this aspect clear as suggested.

P13 L20: I suggest to link here the table S6 and S7 no called in the main text.

We will add this information as suggested.

P13 L27-28: this sentence needs to be followed by some hypotheses for this bi-modal pattern, especially if physical explanations can help to improve the model for biocides release.

We only can speculate about possible explanations. One possibility we will mention is a bi-modal application pattern. However, we don't have any data on that.

P13 L 32: I can't see how two clusters can be derived from the figure 4. Could you clarify this point?

If the residuals were derived from a single population, the marginal distribution of the (transformed) residuals should follow a unimodal distribution. Inspection of Fig. 4 however (most pronounced TBA, DIU or CBZ, but also visible for the others) reveals that the marginal distribution (for concentrations above base flow concentrations) are bimodal: one modus reveals negative values, one modus positive values.

P13 L39: because the GRI is probably less known compared to the NSE criteria, it could be relevant to provide quality thresholds to consider poor, acceptable, good and very good capabilities of the model.

We understand the wish to have guidelines how to judge any reported GRI value. However, we used this metric as a descriptive tool. Whether or not the reported values are deemed satisfactory or not depends on several factors such as the purpose of the model application, the variables modelled (discharge, nutrients, sediments (see e.g. Moriasi, Arnold et al. 2007), micropollutants etc.), and always has a certain level of subjectivity.

One possibility to overcome (partially) this subjective aspect would be a broad comparison across models and case studies - similar to (Moriasi, Arnold et al. 2007) – for model performance on herbicides and biocides. This would provide an empirical basis of how such models can perform under different conditions. This task however, is beyond the scope of this paper. Therefore, we do not provide any ad-hoc recommendation on what should be considered adequate model performance as measured with the GRI.

P14 L3: could you provide in SI Table S9 the unit of RRMSE (%?) Its not clear in SI if it's really in % (very low value if in %). The table S9 is not used to support quality of the model in the main text.

Probably this comment refers to Table S8 because S9 does not contain RRMSE. The RRMSE is defined in Table 3. It is a unitless quantity. We will indicate this in the caption.

We will refer to Table S9 in the main text.

P14 L2: what could be the interest to reproduce the cumulative concentration (observed and simulated sorted) if the dynamic (timing) is not correctly capture?

For water quality assessment, the exact timing is generally not of interest. The relevant questions are more about the degree, frequency or duration of exceedances of water quality standards. For this reasons, the cumulative distribution is of interest.

We will add a sentence for explanation.

P14 L16: how the reader can derive this information? Which tables or figures supports this statement?

We will add the reference to figures S23 and S24.

P14 L24-25: If I well understand, you observed more dissipation than prior estimation. Does it mean that missing dissipation pathways (leaching to deeper groundwater) could be counterbalance in the model by more sorption (model structure error)?

Actually, we observed lower degradation rates (p. 14, L: 24) but due to higher sorption the mass for transport was available for a longer period.

This observation might counterbalance some model errors, for example regarding the timing of herbicide application. If our modelled input was restricted to a too short period this might be compensated by the parameters characterising the fate in the soil.

We may briefly discuss this possibility in the text.

P15 L2-3: do you mean cumulative distribution or chemogram (dynamic evolution of concentration)?

Yes, we refer to the cumulative distribution and will mention this explicitly in the text.

P15 L4: Could you provide hypothesis or compare this behaviour whith other similar approach (calibration in small scale and validation at large range of scale)?

One hypothesis is that the input estimates are more reliable at the larger scale because regional differences and variabilities in local application dates – hence also input uncertainty - are averaged out. We will mention this in the text.

P15 L15: do you have hypothesis for this over-estimation? Could this overestimation in the validation sites be due to an overestimation of the diuron release in the calibration sites owing to agricultural use neglecting?

Input uncertainty is for sure a very plausible reason. However, agricultural use of diuron is hardly the reason because of the small areal fraction of vineyards.

P15 L26: Could you provide figures or tables (in the main text or SI) to help the reader to understand the differences between the 2 routing methods?

We will add a simple scheme to the SI.

Perhaps it could be relevant to sort the results depending of the catchment area to underline the threshold from which the full routing model improve the concentration prediction.

Sorry, we should have referred to Table S10, where we show for which catchment size routing was considered. We will do so in the revised version.

P15 L39-40: you mentioned previously that you did not simulate the biocide due to a lack of the input database for Germany and France.

We keep this sentence for clarity.

P15 L2: missing figure number! Figure 8, I guessfor

Yes, it should have been Fig. 8.

P15 L4: why applications during fall were not considered? Isoproturon is usually applied in October on winter wheat

There are different reasons for not considering the fall application. First, the fall application was not included in any of the calibration studies. Hence, we lack actual data for model calibration. Second, based on expert interviews with a plant protection specialist in Switzerland, we assumed that most of the isoproturon would be used in spring. Thirdly, predicting the timing of fall application is more complex than the spring application, which can reliably derived from temperature sums during the year. Fall applications also depend on the timing of harvest of the previous crop, which depends on the crops and the growth conditions over the entire growing season. For these reasons, we only considered the spring application.

We discussed this on p. 17, L: 22 – 26. We may add a sentence on how to potentially overcome this shortcoming by deriving an stochastic application model based on application data that may get available from national surveys.

P16 L16: Probably not at the larger scale! It's probably a strong hypothesis for which scenarios with and without river processes (degradation and sorption) could be tested with available dataset (DT50 water...)

As mentioned in the text, this hypothesis (or assumption) is very much compound dependent. Compounds that undergo rapid transformation (e.g., due to hydrolysis, photolysis or biological transformation) had definitely to be treated differently.

However, if a compound is chemically stable, the biological degradation may be not as relevant as possibly expected based on lab-based fate studies. Careful analysis of existing OECD lab data on DT50 reveal that these values have little direct relevance on the actual fate in streams unless the interaction with the sediment layer in the experimental systems and the streams are consistent (Shrestha, Junker et al. 2016).

Some of us have recently analysed longitudinal concentration trends for pharmaceuticals along river Rhine in order to test to which degree degradation and sorption can be derived from such data. It



turns out that first, given the input uncertainty the differentiation between conservative and non-conservative behaviour is hardly possible (Honti, Bischoff et al. submitted).

P16 L15: same comment as previously for unpublished paper (possible in HEES?)

See related response above.

P22 L6: Figure 1 : I suggest a modification of the titre: the study area covered the Rhine river upstream the Emmerich discharge gauge (red circle) I suggest also to delineate more clearly the Rhine basin with bold line

We will modify the figure as suggested.

P24 L3: Figure 3: I suggest to express concentration in g/L to be able to better link them to EU drinkable threshold (0.1 g/L) I also suggest to add application dates to see interplay between application and rainfall calendars.

We prefer to stick to the ng/L because this is more reader-friendly in figures where low concentrations are depicted (see Fig. S17 – 22).

We will indicate the application periods for the herbicides in the figures.

For biocides, P/Q is not easily understandable, is it a ratio?

Yes, P/Q is the ratio between precipitation and discharge. The reason for this scale is that i) the simulated load is proportional to precipitation (Eq. 8), and that ii) the concentration results from the load divided by discharge.

The sentence (Line 26-27, page 13) that biocide concentration follows rainfall patterns can not easily be derived from this figure), especially at the end because legend hides the rainfall/discharge dynamic (probably high)

We will provide a figure (possibly in the SI) depicting concentrations as a function of the P/Q ratio.

Finally, on the left you have to remove one ng/L (two times in the figure)

We will correct this mistake.

P16 L25: I agree for herbicides, but it could be remembered that seasonal peaks of biocides were not well represented. This weakness seems not to be discussed in the paper. Which types of processes can explain a seasonal variation of biocides exports from roofs and walls?

We mention the limitations regarding biocide concentrations several times in the text and discuss it briefly in 6.2 (Model limitations, p. 17, L: 21, 26 – 27). We add a sentence suggesting that input uncertainty may be the reason behind.

P16 L27: the authors argue that iWaQa can help to identify potential hotspots in river network. I'm not convince according the way that herbicides loads is calculated. Could the authors more clearly underline the strengths and weaknesses of this model to identify hotspot according conceptual structure and processes taken into account? If I well understand, the inputs are derived from

administrative data and can not explain extreme applications and associated peak of concentration at small scale.

We agree that the spatial resolution of “hotspots” in the river network cannot be better than the spatial resolution of herbicide input data. We will clarify this aspect and point out that the model does not aim at identifying single fields in small catchments but at identifying critical regions: “The spatial resolution of such an analysis however, may be strongly limited by a lack of spatial data on compound use and data on local factors influencing transport. Accordingly, it is expected to be valid at a regional instead of a local scale.”

P16 L35: missing number of figure Table 5 and Figure : : :?

Sorry for the mistake. It should be Fig. 9.

P17 L5: modify the way that the reference is called. I suggest “as discussed in Honti et al. (2017)

We will modify as suggested.

P17 L21 to27: I suggest to order the different elements (1) too high herbicide: : :, (2) seasonal biocide : : : and (3) the lack of an isoproturon application: : : In the following sentences, I only see discussion on the points 3 and 2 and not for the point 1. Reorganize this section.

We will modify as suggested.

P17 L29: “This agrees with the findings from the error models”, I suggest to link here the tables in SI providing range of error.

We will modify as suggested.

P17 L37: missing number for the figure mentioned

Sorry for the mistake. It should be Fig. 10.

P25 Figure 4: even if the calculation is remembered in the caption, it could be relevant to add in the different box of this figure, arrows and captions underlining that if residual > 0 the model underestimated concentration and the opposite if < 0. I also suggest to move this figure in the SI.

We will add this information as suggested and move the figure to the SI as also suggested by Reviewer 2.

P26 L2: I suggest to explain that the catchments are sorted by size (with an arrow and associated surface to better support the sentence (Line 4-5, page 15)

We will modify as suggested.

P28 Figure 7: move to SI

We will add move the figure to the SI as also suggested by Reviewer 2.

P30 Figure 9: move to SI

We will add move the figure to the SI as also suggested by Reviewer 2.