

Revision of the manuscript (point by point explanations)

To all reviewers

Thanks for the detailed and constructive comments. Below we added our initial responses (in red) to each single point and listed how the manuscript has finally be modified (in green, line numbers refer to the revised version).

We also appended at the end the revised manuscript in the track-change mode such that one can easily identify where the paper was revised.

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.

We have first modified Fig. 2 such as to clearly show the spatial arrangement of the calibration and validation catchments within the river network. Additionally, we have added a schematic representation in the SI that depicts also the size relations between the catchments (Fig. S4).

The results are clearer for herbicides than for biocides. The reader discovered different hypotheses that reduce progressively the extent of the biocides loads and 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.

No changes were made.

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.

We corrected the terminology throughout the manuscript incl. the SI.

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

Will be corrected.

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.

We have modified the text to i) highlight the novelty more and to be explicit about the temporal resolution.

"Here we present a model that covers major urban and agricultural sources for pesticides in streams that can be applied to large water basins, provides high spatial and temporal resolution (hourly to daily) and is still parsimonious. It is similar to the iWaQa model approach in (Honti et al., 2017) but adapted for large basins by including an explicit routing component by coupling it to the AQUASIM model. It differs from many other model concepts in that it does not include a rainfall-runoff module but directly links agricultural pesticide losses in a novel way to measured discharge and urban biocide losses directly to precipitation." (L. 94 - 99).

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.

Done.

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.

We added two additional figures to the Appendix that show first how output from the substance transfer module from each catchment is routed with the two routing approaches and second the spatial arrangement of the study catchments (Fig. S3, S8).

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.

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."

We added the sentence as mentioned above. (L. 175)

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.

No changes were made.

P6 L23: no rainfall dependency?

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

No changes were made.

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.

No changes were made.

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

We will mention the coupling with Aquasim in both parts.

Done (L. 28, 97).

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

We will skip this reference and the one on page 16

The reference was skipped.

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.

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.

Done. See explanation above (Fig. S3, S8).

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

We will add the respective information.

We added “(804 out 931 or 86 % of the available gauging stations).” (L. 268)

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.” .

We added the explanation as mentioned above. (L. 270 – 271)

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.”

We added the explanation as mentioned above. (L. 274 – 277).

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°C/m altitude increase”.)

Done. (L. 301 – 302)

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).

We added the following text:

“Because there was no data set available reflecting the most recent situation, we checked whether there have been major shifts in agricultural land use with the spatially lumped data on the 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 mostly little 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 (see Fig. S4). For Switzerland, more recent land use (2004 – 2009) and crop statistics (2010) were available and used.” (L. 308 – 314)

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.

We added this information in Appendix A4, Table S1.

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 km²) used for model calibration. In these cases however, we rely on specific and detailed land use data for the specific year.

No changes were made with regard to crop rotations

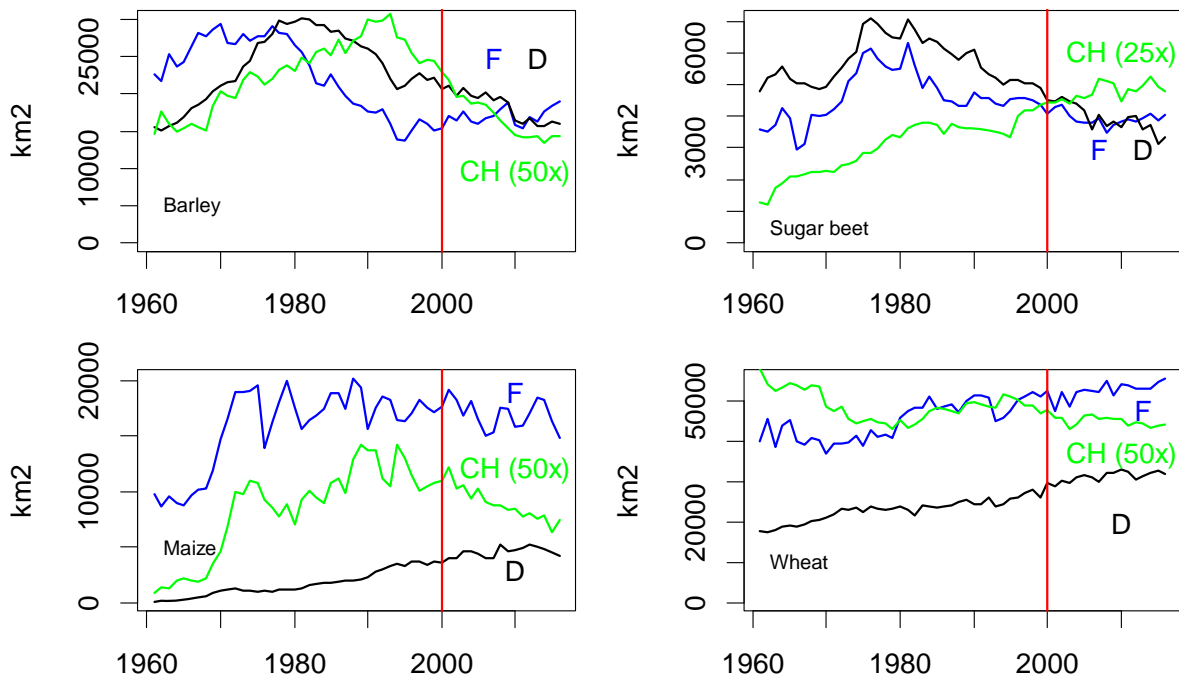


Fig. 1: Temporal evolution of cropping area for four major crops relevant for the modelled herbicide use. The red line indicates the reference year for the spatial cropping patterns. Data source: <http://www.fao.org/faostat/en/#data>.

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.

No changes were made.

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 added the sentence “The resulting spatial distribution of estimated input is depicted in Appendix A6, Fig. S5.”

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.

We added “Because of the lack of spatially distributed biocide use data, the spatial distributions of CBZ and DIU are identical (see Appendix A6, Fig. S6). (L. 376 – 377)

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

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

No changes were made.

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.

We think that the explanation on the error-scaling function provided below is sufficient at that point.

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.

We added the sentence as described above. L. 396 – 398.

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).

We modified the sentence: “The likelihood function used in this study is based on the assumption that Box-Cox transformed (Box and Cox, 1964) time series of concentration data C lead to independent and identically distributed normal errors as described in Honti et al., (2017). The corresponding likelihood function is as follows:....” . (L. 405 – 407)

P11 L10: studies are mentioned to assess the prior distribution of ε , 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).

Done. L. 424, Tab. S3.

P11 L28: I suggest to mention the table S4 and S5 (Appendix A8) after “of the priors for ε and β ”

We will add this information.

Done. L. 442.

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

We will modify the text accordingly.

Done. L. 476.

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.

We modified the respective sentences to:

“Subsequently, it was validated for herbicides and biocides in Switzerland for different years on 12 catchments of much larger size (31 –36’000 km²) and for herbicides for the entire Rhine basin upstream of the Dutch-German border (160’000 km²) without any modification.” L. 31 – 34.

In the Introduction we modified the last sentence: “Due to lack of data, the biocide part was only tested within Switzerland.” L. 109 – 110.

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.

Done. L. 505.

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.

We added the following sentence:

“A possible reason for this temporal pattern is a seasonal application pattern of biocide application. However, there is no data available for testing this hypothesis.” L. 518 – 519.

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.

Fig. 4 was moved to the SI (new: Fig. S11). No further changes were made.

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.

No changes were made.

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.

Done.

We will refer to Table S9 in the main text.

Done. L. 509.

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.

We modified the sentence by inserting:

– this generally relevant for water quality assessment - (L. 531)

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.

These figures are mentioned in the text. Now Fig. S27 – S28, see L. 548.

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.

We added a sentence:

“However, stronger sorption in the model could also compensate for pesticide applications that were missed by the model.” L. 556 – 557.

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.

We inserted “(cumulative)” L. 576.

P15 L4: Could you provide hypothesis or compare this behaviour with 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.

We added a sentence: “This might be explained by averaging out regional differences and variabilities in local application dates – hence also input uncertainty - across larger scales. L. 579 – 580.

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.

No changes were made.

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.

We added Fig. S3 to that end.

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.

Done. L. 603.

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.

No changes were made.

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

Yes, it should have been Fig. 8.

Done. L. 586.

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.

We added a sentence: “In the future, this deficit may be overcome by deriving a stochastic application model based on application data obtained from national surveys.” L. 688.

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).

No changes were made.

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

See related response above.

Done.

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.

We modified Fig. 1 as suggested.

Title: Because the title does not imply that the simulations covered the entire basin, we suggest to keep the title for the sake of brevity.

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.

We indicated the application periods for the three herbicides in Fig. 3.

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.

No changes were made.

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.

We modified Fig. 3 such that the high values at the end of the year can be properly seen. Additionally, we added Fig. S10.

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

We will correct this mistake.

Done.

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.

We added “... while that of biocides was not well reflected (see below)” L. 641, and referred to the limited model performance for biocides at several locations (L. 532, 575, 659, 689 – 692)

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.”

We added the explanation as mentioned above. L. 642 - 644.

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

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

Done. Now Fig. S34, see L. 653.

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

We will modify as suggested.

Done. L. 663.

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.

We have extended this paragraph to the following version:

“The first problem of herbicide background concentration levels would require a more explicit modelling of the long-term fate of these compounds in the coupled unsaturated and saturated zone. To keep such a model parsimonious one had to test whether these background concentrations could be empirically linked to some simple catchment characteristics. Second, the herbicide application in fall for example is much more difficult to predict compared to the spring application because it not only depends on a single variable such as the temperature sum over the year but it is also influenced by all the climatic variables determining the time of cropping of the previous crop and potential intercropping. In the future, this deficit may be overcome by deriving an stochastic application model based on application data obtained from available from national surveys. Regarding the third aspect, the seasonal biocide patterns, we lack any information about biocide use on buildings that could explain the observed seasonality. Targeted surveys on actual use across the year might be a solution.” L. 682 – 692.

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.

We refer to Appendix, Fig. S27, S28, L. 694. .

P17 L37: missing number for the figure mentioned

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

Corrected. Now Fig. 7, L. 702.

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.

Done. See Fig. S11.

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.

We indicated now the size of the catchments but didn't add an arrow to not overload the figure. (Fig. 4).

P28 Figure 7: move to SI

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

Done. Now Fig. S33.

P30 Figure 9: move to SI

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

Done. Now Fig. S34.

Anonymous Referee #2

Do the authors give proper credit to related work and clearly indicate their own new/original contribution? *NO, literature on state-of-the-art is incomplete*

We will carefully update the cited references and pay due attention to the suggestions by the reviewer. Thanks for those hints.

Done. L. 69 – 72.

Is the overall presentation well structured and clear? *MOSTLY, some figures are not*

We will i) modify some of the figures according to the comments and suggestions by the reviewers (see also response to Reviewer 1).

Done. See above.

Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated? *Substantially clarify figures and/or reduce numbers of figures*

See previous comment. We will also move Fig. 4, 7, and 9 to the SI as was also suggested by Reviewer 1.

Done. See above.

Specific comments

Literature with respect to other related model work in this field is not sufficiently reviewed. References are sometimes unfortunate, e.g. MONERIS is cited not with original papers but with a side application and reference to GREAT-ER is outdated (see Kehrein et al., 2015). The EU approach (FOCUS model suite) is neglected as is the DRIPS model (Röpke et al., 2004).

We will carefully update the cited references and pay due attention to the suggestions by the reviewer. Thanks for those hints. Additional references include: (Behrendt, Kornmilch et al. 2002, Röpke, Bach et al. 2004, Kehrein, Berlekamp et al. 2015, Steffens, Jarvis et al. 2015, Villamizar and Brown 2017).

We have included additional references as follows:

“Some of these models (e.g., SWAT (Arnold et al., 2011), MONERIS (Behrendt et al., 2002), GREAT-ER (Berlekamp et al., 2007;Kehrein et al., 2015)) or MACRO (Steffens et al., 2015;Larsbo et al., 2005) have been widely used, many others have been developed and used in specific research contexts (e.g, ZIN-AgriTra (Gassmann et al., 2013), SPIDER (Renaud et al., 2008;Villamizar and Brown, 2017) or DRIPS (Röpke et al., 2004)).” L. 69 – 72.

The assumption of general presence of persistent pesticides in baseflow and groundwater (page 6) is somehow terrifying to me, since prevention of elevated concentration levels in groundwater is the main driver for research in this field. Background concentrations determined in the calibration step need to be explicitly compared to the large number of data available. Fortunately, many of these data points are small and below critical limits. These data allow for defining an upper bound for the calibrated background concentrations.

The occurrence of (low) levels of background concentrations of some herbicides in groundwater is a well established fact (Vonberg, Vanderborght et al. 2014, Hakoun, Orban et al. 2017) and supported by our data sets (Gomides Freitas, Singer et al. 2008). An example for this persistence is provided with daily atrazine concentrations in river Rhine at Basel. Even years after the ban of this compound it is ubiquitously found.

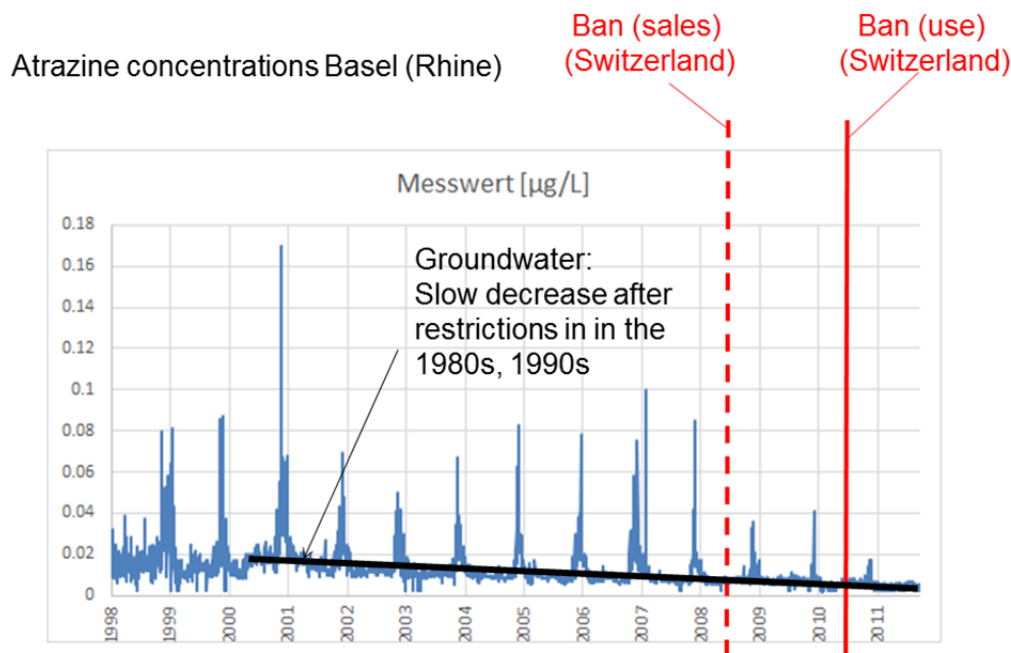


Fig. 2: Temporal evolution of atrazine concentrations in river Rhine at Basel. Data source: IRMS.

The background concentration reflects the concentration in the groundwater which is an important contributor of water and substance during base flow conditions. The concentration in the groundwater is assumed to be proportional to the intensity of herbicide use in the area. Hence for simulations the calibrated background concentrations was proportional to areal fraction of the respective crop in a subcatchment. Thus a lower background concentration was introduced at subcatchments with a small share of agricultural area and vice versa a higher background concentration was selected for subcatchments with intensive agriculture land use (as mentioned in Table 1).

The means in the prior distributions of the background concentrations for the calibration were estimated from the values in Table S4.

No changes were made.

Many German cities situated in the Rhine catchment use combined sewer systems, whereby a significant fraction of urban surface runoff is directed not instantaneously into rivers (page 6), but to the nearest sewage treatment plant. While the substance-specific load reduction during wastewater treatment is included in the loss rate, it should be discussed whether retardation of transport into the receiving waters for the hydraulic retention time in the sewer system and the wastewater treatment plant could have an effect on the timing of the biocide concentration peaks in the simulations.

The short-term timing is less of an issue for the biocide simulations; the seasonal deviations are more striking. Additionally, the lag time for transport through the urban sewer system causes not a severe problem due to the daily time step used for our simulations. Nevertheless, the point is well taken and we will add a comment in the text.

We added a sentence:

“The assumption of instantaneous transfer to the stream may cause some timing errors if compounds have residence times that are longer than the model time step (e.g. in wastewater treatment plants) but see the findings on routing effects in sec. 5.2. “ L. 229 - 231.

I think that some figures in the main text are not important enough to justify their presence. Moreover, the meaning of some figures is hardly recognizable from the discussion. I strongly encourage the authors to shift some of the figures to the supporting information. In the main text, I would focus the discussion on the most important points supported by meaningful and easily understandable figures with clearly labelled axes.

See previous comment. We will move Fig. 4, 7, and 9 to the SI as was also suggested by Reviewer 1.

Done.

- Figure 4: I could not find a description of the transformation step used to generate the data displayed here. Both axes are displayed in dimensionless scale, but I have no idea what that means. It remains also unclear, what the dashed line represents and why the Figure points to systematic deviations between observed and modelled concentrations (page 13, line 31).
The reviewer is correct in that the figure was poorly labelled. As described on p. 10, L. 30, the calibration was performed on the Box-Cox-transformed data (Eq. 14, p, 10, L. 36 – p. 11, L: 2). The dashed line represents the regression of the residuals versus the observed concentration (in the transformed space). Without systematic deviations, the regression slope should not deviate from zero.
We will add the necessary information in the figure and captions. Additionally, it will be moved to the SI. In the main text, we will better explain how to identify the systematic deviations.

We moved the figure to the SI and improved its description.

- Figure 5: It is not possible to extract the numbers for the GRI values given in the text (page 13/14) from the graph. For me, the range shown in the graph is larger with values up to 5 for the herbicides. Meaning of vertical lines is unclear and the error bars (grey) unreadable. Colour of second section (C-T, C-S) is not identifiable and it is confusing that all categories appear twice on the category axis.
We will better label the figure and use additional colours to improve its readability. Furthermore, we make the link between statement in the text and the figure more implicit since it seems to be confusing.

Done. Now Fig. 4.

- Figure 10: Scaling of the axes is unclear (log-scale of what?)
The axes represent the average density of data series in space (x-axis) and time (y-axis) across the model domain.

- The x axis represents the number of available data points X across the entire study area (e.g., how many discharge gauging stations, in log scale)
- The y axis gives the number of available data points Y per unit area per decade (, e.g., how many rainfall data for one location in space for a ten year period, in log scale)
- The radius of the circles is proportional to the total number of data points during a unit period of time for the entire study area (product X times Y, in log scale). We will modify the circles such that the circle area is proportional to the $\log(X Y)$.

We have completely re-drawn this figure and hope that it is now easy to understand. Now Fig. 7.

Model performance for biocides is poorer than for pesticides (page 14, line 2), which should be discussed in the light of the different number of calibrated parameters for herbicides (seven/nine with error function) and biocides (just one). Either the biocide model needs more detailed process descriptions or the herbicide model is over-parameterized mutually levelling off uncertainties introduced by the parameters.

The results suggest that the biocide model is overly simplistic because it cannot produce the observed seasonality. However, it is difficult to improve upon this aspect first because the underlying cause is not understood and second because of poor input data that preclude a more reliable simulation. We will add a sentence in the discussion to elaborate more on this issue beyond we already have on p. 17, L. 34 – 38.

We added the sentence:

“Targeted surveys on actual use across the year might be a solution. Better input data could then allow to study further structural deficits of this very simple biocide model in more detail.” L. 690 – 692.

Minor comments

- pathways, wastewater (page 2)
Will be corrected.

Done. L. 12.

- ... where $M(t)$ [g d⁻¹] is the mass ratio ... (page 5)
We will reword to “ is the rate of mass applied...”

Done. L: 174.

- ... temperature sum models is used (page 5)
Will be corrected.

Done. L. 180.

- ... of the pronounced concentrations peaks (page 15)
Will be corrected.

Done. L. 604.

- Number of Figure missing (page 16, line 2)
The correct figure number will be inserted.

Done. Fig. 6, L. 616.

- ... for the model compounds considered in this paper (page 16)
Will be corrected.

Done. L. 632.

- ...micropollutants from points sources (page 16)
Will be corrected.

Done. L. 640.

- ...purely hydrological models can accomplish (page 16)
Will be corrected.

Done. L. 656.

- Number of Figure missing (page 17, last but one line)
The correct figure number will be inserted.

Done. L. 702.

Anonymous Referee #3

Received and published: 6 March 2018

The authors present an extensive modeling study on pesticide transport in the Rhine basin. The paper including the supplement is very long (as is the river Rhine). It builds on quite an impressive amount of work and is well written, although not everything is explained in full detail. The latter is probably unavoidable with such kind of studies. I appreciate that the authors provide there input data and model code.

The basic idea is to set up a simple, parsimonious model, to calibrate it with detailed data (from Switzerland, where a number of very good, detailed studies were carried out in the past) and predict transport on large scale. The approach is actually quite successful. Modeled concentrations partly deviate from measured concentrations by up to a factor of 5 or so, but we should also not forget that the concentrations are low.

It is clear that the model can be improved when better input data become available (if ever).

I have only one major point to make. I refer to the last paragraph of section 5.1. The authors used a (lumped) modeling approach in which sorbed phase and solution phase concentration are expressed in total masses in the system (watershed). The model was parameterized with laboratory data from a study by Freitas et al. 2008 (reference missing), but it is not explained how.

Was the different soil to solution ratio in the lab considered? If the model is directly applied to the lab data (M_s =sorbed, M_w =dissolved), it would yield too low partitioning coefficients, what is indeed what the authors observed (section 5.1, last paragraph). This is so because, as a physicochemical fact, the sorption isotherm is independent of the soil to solution ratio as has been shown many times in the literature. In the lab, the mass in solution is much higher (by a factor of 5 or so, depending on the soil to solution ratio used). Taking it differently, sorption isotherms measured in the lab cannot be applied to a system (watershed) without making an assumption about the soil to solution ratio in it.

The reference was included in the reference list but as Gomides Freitas et al., 2008. We will correct this mistake and make the referencing consistent.

Done. L. 445.

The data stem from a field study (p. 11, L. 29 – 30). Pore water concentrations from soil samples were obtained upon extraction of 20 g of soil with 40 mL 0.01 M CaCl_2 solution. Therefore, the reviewer is right in that the soil to solution ratio does not correspond to the conditions in field soils. How this fact affects the partitioning between the aqueous and the solid phase depends on the sorption isotherm. With a linear isotherm, the distribution coefficient K_d is constant irrespective of the soil-solution ratio. With non-linear isotherms this is not the case. Our field data indicate non-linear sorption behaviour that can be described with a Freundlich isotherm.

Based on the experimental data, we can calculate how the (apparent) distribution coefficient K_d is expected to change upon a modification of the soil-solution ratio. To this end, we used the S-metolachlor data from (Gomides Freitas 2005) with a Freundlich exponent n of 0.7 and a Freundlich coefficient of 17 and calculated the distribution coefficient between solid and the aqueous phase for the lab conditions (see above, $\rho = 0.33$, $\theta = 0.67$) and for realistic field conditions ($\rho = 1.00$, $\theta = 0.25$). The resulting ratios between the lab and field distribution coefficients are depicted in Fig. 3 across a wide range of total concentrations (10 to 2500 microgram/kg soil).

These calculations first demonstrate that the Reviewer is correct in that the lab data underestimate sorption. Second, the data also reveal that the degree of underestimation is rather limited such that it won't affect our findings substantially. Nevertheless, we will add a comment stating that the different soil-solution ratios between the measuring conditions and the field situation will have contributed to the underestimation of partitioning.

We added a sentence as follows:

“For sorption, this could be explained to some degree by different soil-water ratios of undisturbed soils and the conditions during the experimental procedure from which the priors were derived (Gomides Freitas, 2005). However, stronger sorption in the model could also compensate for pesticide applications that were missed by the model.” L. 555 – 557.

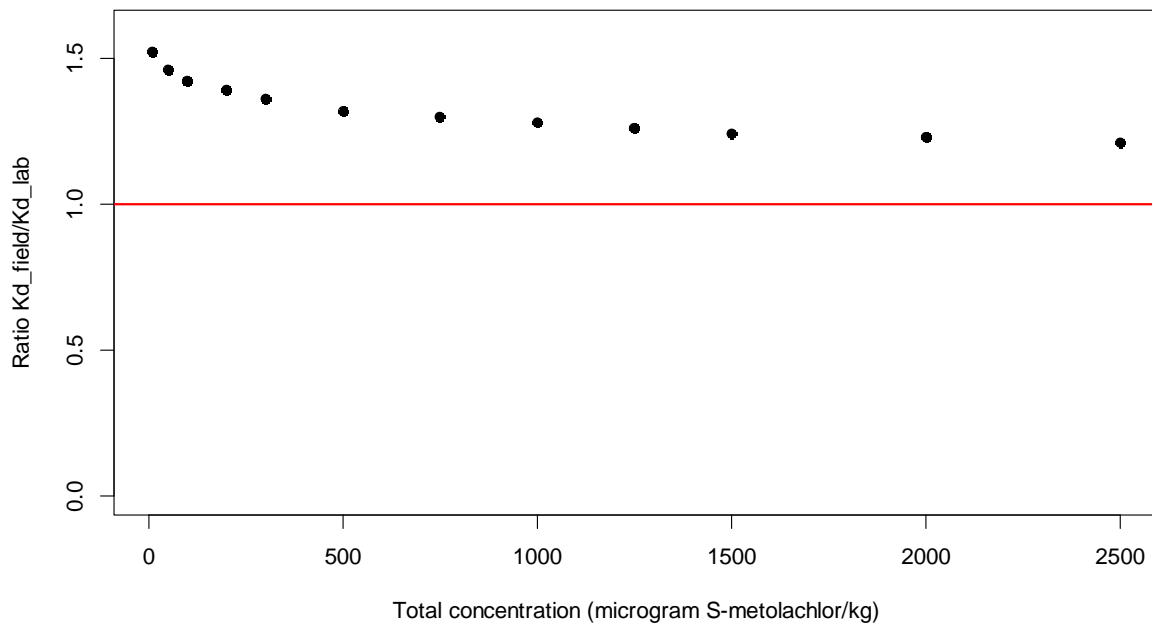


Fig. 3: Ratio between distribution coefficients in the field and in the lab for S-metolachlor based on the data reported in (Gomides Freitas 2005). The red line indicates identity between the lab and the field data.

Unfortunately, my print-out version did not contain page numbers.

page 1, line 17 to what degree

Will be corrected.

Done. L. 17.

At some locations in the text, references to figures lack the number (figure [empty]).

Will be corrected (see response above).

Done.

eq 1: $M(t)$ is undefined (cf. eq 8)

This mistake will be corrected.

Done. L. 167.

eq 8: in eq 2 and 3, M_a was defined as the mass applied to the catchment the mass present in the catchment should be M as in the LHS of eq 1

Will be corrected.

Done. L. 228.

Figure 4: how were the residuals transformed?

All data are Box-Cox transformed, see response to Reviewer 1 on the same issue.

Done.

Figure 5 Overview of

Will be corrected.

Done. Now Fig. 4.

Figure 6: I did not get the rational behind the fold difference. This should be explained in the text and in the figure caption.

In the caption, we will refer to the description in Tab. 3 and we will add additional explanations in the text saying that this metric allows to quantify the deviation between observations and simulations irrespective of whether observations were under- or over estimated.

We have inserted

“– indicating by which factor observations were over- or underestimated –” L. 585 – 586.

Figure 7: What is the violin about? Is the envelope a standard error? Explain in the text and in the figure caption.

Violin plots are a combination of a boxplot and a kernel density plot. We will add a further explanation in text and captions and refer to the original publication introducing this kind of plot (Hintze and Nelson 1998).

We have moved Fig. 7 to the SI and added “– combination of a boxplot and a kernel density plot (Hintze and Nelson, 1998) –” in the figure caption (now Fig. S33).

Figure 9: Explain the envelope?

Violin plots are a combination of a boxplot and a kernel density plot. We will add a further explanation in text and captions and refer to the original publication introducing this kind of plot (Hintze and Nelson 1998).

We have moved Fig. 9 to the SI and added “– combination of a boxplot and a kernel density plot (Hintze and Nelson, 1998) –” in the figure caption (now Fig. S34).

Figure 10: I have no clue what this figure is about. Explain the axes, assign units. Explain properly or delete the whole thing.

We admit that the figure was poorly explained. We will add the information provided in the corresponding response to Reviewer 2 (see above).

We have completely re-drawn this figure (now Fig. 7).

Table 1: Did you properly explain the scaling factor ? I think no.

The scaling factor is introduced and explained on p. 10, L. 20 – 26. In accordance with the reply to reviewer #1 on the same issue, we will add further explanation

We added “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.” L. 396 – 398.

Table 2: Explain the abbreviations (like NADUF).

We will add these explanation (NADUF: NADUF – National long-term surveillance of Swiss rivers, IRMS: International Rhine Monitoring Station, NAWA SPEZ: National Surface Water Quality Monitoring Programme) and make sure they are used in consistent forms across the manuscript.

We explained the abbreviations upon first mentioning and also added the explanations to the respective figure and table captions.

Table 3: Explain the symbols in the caption.

We explain the symbols

Done.

Table 4: Explain abbreviations.

We will add these explanations.

Done.

Table 5: C and V should be defined. Atrazine

We will explain these symbols (C = calibration, V = validation). It is unclear what the reviewer wanted to say about atrazine.

Done.

***Summing up, the tables and figures should be self-explaining.

We agree and improve according to the responses above.

Done.

References:

- Behrendt, H., M. Kornmilch, D. Opitz, O. Schmoll and G. Scholz (2002). "Estimation of the nutrient inputs into river systems - experiences from German rivers." Journal of Material Cycles and Waste Management **3**(1-3): 107-117.
- Box, G. E. P. and D. R. Cox (1964). "An analysis of transformations." Journal of the Royal Statistical Society, Series B. **26** 211–252.
- Gomides Freitas, L. (2005). Herbicide losses to surface waters in a small agricultural catchment Diss ETH 16076, Swiss Federal Institute of Technology.
- Gomides Freitas, L., H. Singer, S. R. Müller, R. Schwarzenbach and C. Stamm (2008). "Source area effects on herbicide losses to surface waters - A case study in the Swiss Plateau." Agriculture, Ecosystems & Environment **128**: 177 - 184, doi:110.1016/j.agee.2008.1006.1014.
- Hakoun, V., P. Orban, A. Dassargues and S. Brouyère (2017). "Factors controlling spatial and temporal patterns of multiple pesticide compounds in groundwater (Hesbaye chalk aquifer, Belgium)." Environmental Pollution **223**: 185-199.
- Hintze, J. L. and R. D. Nelson (1998). "Violin Plots: A box plot-density trace synergism." The American Statistician **52**: 181-184.
- Hirsch, R. M. (1979). "An evaluation of some record reconstruction techniques " Water Resources Research **15**: 1781 - 1790.
- Honti, M., F. Bischoff, A. Moser, C. Stamm and K. Fenner (submitted). "Micropollutant degradation in rivers: Suitability of field monitoring and regulatory data."
- Kehrein, N., J. Berlekamp and J. Klasmeier (2015). "Modeling the fate of down-the-drain chemicals in whole watersheds: New version of the GREAT-ER software." Environmental Modelling & Software **64**: 1-8.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel and T. L. Veith (2007). "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations." Transactions of the American Society of Agricultural and Biological Engineers **50**: 885 - 900.

Röpke, B., M. Bach and H. G. Frede (2004). "DRIPS - A DSS for estimating the input quantity of pesticides for German river basins." Environmental Modelling and Software **19**(11): 1021-1028.

Shrestha, P., T. Junker, K. Fenner, S. Hahn, M. Honti, R. Bakkour, C. Diaz and D. Hennecke (2016). "Simulation studies to explore biodegradation in water-sediment systems: from OECD 308 to OECD 309. ." Environmental Science and Technology **50**: 6856-6864.

Steffens, K., N. Jarvis, E. Lewan, B. Lindström, J. Kreuger, E. Kjellström and J. Moeys (2015). "Direct and indirect effects of climate change on herbicide leaching - A regional scale assessment in Sweden." Science of the Total Environment **514**: 239-249.

Villamizar, M. L. and C. D. Brown (2017). "A modelling framework to simulate river flow and pesticide loss via preferential flow at the catchment scale." Catena **149**: 120 - 130.

Vonberg, D., J. Vanderborght, N. Cremer, T. Pütz, M. Herbst and H. Vereecken (2014). "20 years of long-term atrazine monitoring in a shallow aquifer in western Germany." Water Research **50**: 294-306.

Modelling biocide and herbicide concentrations in catchments of the Rhine basin

Andreas Moser¹, Devon Wemyss^{1#}, Ruth Scheidegger¹, Fabrizio Fenicia¹, Mark Honti², Christian Stamm¹

¹Eawag – Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, 8600, Switzerland

²MTA-BME Water Research Group, Hungarian Academy of Sciences, Budapest, 1111, Hungary

[#]current address: ZHAW School of Management and Law, Winterthur, 8400, Switzerland

Correspondence to: Christian Stamm (christian.stamm@eawag.ch)

10 **Abstract.** Impairment of water quality by organic micropollutants such as pesticides, pharmaceuticals or household
chemicals is a problem in many catchments worldwide. These chemicals originate from different urban and agricultural
usages and are transferred to surface waters from point or diffuse sources by a number of transport pathways. The
quantification of this form of pollution in streams is challenging and especially demanding for diffuse pollution due to the
high spatio-temporal concentration dynamics, which requires large sampling and analytical efforts to obtain representative
15 data on the actual water quality.

Models can also be used to predict ~~information to which-what~~ degree streams are affected by these pollutants. However,
spatially distributed modelling of water quality is challenging for a number of reasons. Key issues are the lack of such
models that incorporate both urban and agricultural sources of organic micropollutants, the large number of parameters to be
20 estimated for many available water quality models, and the difficulty to transfer parameter estimates from calibration sites to
areas where predictions are needed.

To overcome these difficulties, we used the parsimonious *iWaQa* model that simulates herbicide transport from agricultural
fields and diffuse biocide losses from urban areas (mainly façades and roof materials) and tested its predictive capabilities in
25 the Rhine River basin. The model only requires between one and eight global model parameters per compound that need to
be calibrated. Most of the data requirements relate to spatially distributed land use and comprehensive time series of
precipitation, air temperature and spatial data on discharge. For larger catchments, routing was explicitly considered by
coupling the *iWaQa* to the AQUASIM model.

30 The model was calibrated with data sets from three different small catchments (0.5 – 24.6 km²) for three agricultural
herbicides (isoproturon, S-metolachlor, terbuthylazine) and two urban biocides (carbendazim, diuron). Subsequently, it was
validated for herbicides and biocides in Switzerland for different years on 12 catchments of much larger size (31 – ~~160'000~~
35'899 km²) and for herbicides for the entire Rhine basin upstream of the Dutch-German border (160'000 km²) without any
modification. For most compound-catchment combinations, the model predictions revealed a satisfactory correlation
35 (median r²: 0.5) with the observations ~~and~~. The peak concentrations mostly predicted within a factor of two to four
(median: 2.1 fold difference for herbicides and 3.2 for biocides respectively). The seasonality of the peak concentration was
also well simulated, the predictions of the actual timing of peak concentrations however, was generally poor.

Limited spatio-temporal data, first on the use of the selected pesticides and second on their concentrations in the river
40 network, restrict the possibilities to scrutinise model performance. Nevertheless, the results strongly suggest that input data
and model structure are major sources of predictive uncertainty. The latter is for example seen in background concentrations

Formatted: Font: Italic

that are systematically overestimated in certain regions, which is most probably linked to the modelled coupling of background concentrations to land use intensity.

45 Despite these limitations the findings indicate that key drivers and processes are reasonably well approximated by the model and that such a simple model that includes land use as a proxy for compound use, weather data for the timing of herbicide applications and discharge or precipitation as drivers for transport is sufficient to predict timing and level of peak concentrations within a factor of two to three in a spatially distributed manner at the scale of large river basins.

1 Introduction

50 Mankind uses thousands of synthetic chemicals for many different purposes in households, industries or agriculture (Schwarzenbach et al., 2006;Bernhardt et al., 2017). Many of these compounds reach water bodies during some stage of their life cycle. Accordingly, the impairment of water quality caused by substances such as pharmaceuticals, household chemicals or pesticides is a problem of many catchments worldwide. From an ecological point of view, pesticides are often of special concern because they have been designed to harm a wide range of organisms.

55

Pesticides are used for different purposes. In agriculture, they are used to protect crops from weeds, pests or diseases. However, the same compounds may be also used to fight unwanted organisms on materials such as roofs, façades or ships.

Depending on where pesticides are used, they may reach water bodies via different path-ways. Although pesticides may be ecotoxicologically relevant chemicals even in treated waste water discharged from point sources (Munz et al., 2017;Müller et al., 2002) diffuse pollution is often dominant for these compounds (Moschet et al., 2014). The quantification of this form of pollution in streams is challenging due to the high spatio-temporal concentration dynamics, which requires large sampling and analytical efforts (e.g., Wittmer et al., 2010;Leu et al., 2004a).

65 As a consequence, the water quality status of many water bodies is not quantified sufficiently for properly addressing management and research questions that require a proper understanding about spatio-temporal patterns of pesticides occurring in streams. There may be deficits with regard to the spatial or temporal coverage of data as well as coverage of all chemicals of interest (Moschet et al., 2014).

Spatially (semi-)distributed models can potentially fill such gaps and have been developed and used for decades to do so (Borah and Bera, 2004). Some of these models (e.g., SWAT (Arnold et al., 2011), MONERIS (Behrendt et al., 2002), GREAT-ER ~~or MONERIS~~ (Berlekamp et al., 2007;Kehrein et al., 2015)) or MACRO (Steffens et al., 2015;Larsbo et al., 2005) have been widely used, many others have been developed and used in specific research contexts (e.g, ZIN-AgriTra (Gassmann et al., 2013), SPIDER (Renaud et al., 2008;Villamizar and Brown, 2017) or DRIPS (Röpke et al., 2004)). One of the challenges related to modelling diffuse pesticide losses is the necessity to cover all relevant sources and flow paths. Many models for example, do not simulate urban and agricultural processes with the same level of detail. This may pose a serious problem in regions that are characterized by a mixed land use of urban and agricultural areas such as in many parts of densely populated Central Europe.

80 Models differ widely in the degree to which they aim to represent explicitly the relevant processes. On the one hand, so-called physically based models try to describe them with equations in such a way that the model parameters should have a real physical, chemical, or biological meaning independent of the model application with the goal to provide causal system understanding (Bossel, 1994;Beck, 1987). Generally, running such highly parameterized models comes with a huge data demand, and – as this demand usually cannot be covered – many model parameters cannot be estimated from independent

85 observations. In the end, this leads to either the use of potentially unrealistic parameter values or calibration, the latter facing the problem that many of the parameter values cannot be properly identified possibly inducing large uncertainties during a validation or prediction phase (Beck, 1987; Brun et al., 2001).

90 On the other hand, more conceptual, parsimonious models try to cope with the lack of (spatially distributed) data by dramatically reducing the number of parameters. This comes at the cost that model parameters may lose their direct physical or chemical interpretation. Such parsimonious models basically assume that essential aspects of the response of a complex (real) system can be represented by some rather simple mathematical descriptions that incorporate the effects of major external drivers, such as precipitation. Such types of models are frequently used in hydrology for simulating discharge (e.g., Beven and Kirkby, 1979), there are also applications to water quality simulations (Hahn et al., 2013; Jackson-Blake et al., 2017) but only few models for simulating pesticide transport to surface waters (Honti et al., 2017).

95 Here we present a model that covers major urban and agricultural sources for pesticides in streams that can be applied to large water basins, provides high spatial and temporal resolution ([hourly to daily](#)) and is still parsimonious. It is similar to the *iWaQa* model approach in (Honti et al., 2017) but adapted for large basins by including an explicit routing component [by coupling it to the AQUASIM model](#). It differs from many other model concepts in that it does not include a rainfall-runoff module but [directly](#) links agricultural pesticide losses [in a novel way](#) to measured discharge and urban biocide losses directly to precipitation.

Specifically, the paper has the following objectives:

1. Description of the model concepts and their implementation.
- 105 2. Calibration of the model on selected small catchments and selected pesticides representing agricultural herbicides and urban biocides
3. Evaluating the performance of the calibrated model with blind predictions on a large set of validation catchments. This step includes a pronounced spatial upscaling of the model by three to four orders of magnitude.

110 We have used the Rhine basin [upstream of Emmerich \(see Fig. 1\)](#) as a case study to investigate these questions. [Due to lack of data, the biocide part was only tested within Switzerland.](#)

2 Study area

The study is carried out in the Rhine basin upstream of the gauging station Emmerich am Rhein (Germany; see Fig. 1). We limited the analysis to this part of the basin because the model structure does not cover complex, strongly managed flow regimes as prevalent in the Dutch part of the basin. Even with these restrictions, the study area is one of the largest drainage basins in Europe with an area of 160'000 km² covering land of eight countries, mainly from Switzerland, Germany, France and Luxembourg. The total length of the river network is 63'080 km and is divided into more than 30'000 catchments according to the CCM River and Catchment database for Europe, version 2 (CCM2) from Vogt et al. (2007).

Altitude ranges from above 4200 m.a.s.l. in the Bernese Alps in the south to 17 m.a.s.l. at Emmerich in the north. Accordingly, the hydrological regime varies strongly across the basin. [The discharge regime](#) in the southern part of the basin is largely influenced by snow accumulation and melt. As a consequence, most southern rivers are of pluvio-nival type with low water periods during winter and flood events occurring mainly in summer. In contrast, sub-basins further north (Neckar, Main, Moselle, Ruhr, etc) are characterized by a pluvial regime with winter floods and low water levels in summer.

Similarly, temperature regimes show important differences, which may be reflected in shifts in phenology of crops and hence in application periods of agricultural pesticides.

The basin is densely populated (290 inhabitants km⁻² in the study area) with strong regional differences. Arable cropping is an important land use in large parts of the basin. More details on specific crops and their spatial distribution are presented in the Supplementary Material (Fig. S3S5).

The Rhine River is heavily used by hydropower plants upstream of Iffezheim along the main channel and main tributaries. However, the effects on travel times is rather moderate on the streams of interest and was neglected during the routing calculations.

3 Model description

3.1 Spatial discretization and model structure

The river basin is discretized into subcatchments based on the CCM2 database. To reduce the number of subcatchments and ensure a reasonable minimum size, CCM2 catchments smaller than 2 km² were merged with the next downstream catchment. The resulting 18'240 subcatchments with an average area of 8.8 km² are the primary computational units of the model. Further details on the spatial representation are provided in Appendix A2 of the Supplementary Material.

The model consists of two principal components. The first component – the *substance transfer module* – simulates the transfer of the pesticides from their point of use (e.g., the fields to which herbicides are applied) to the outlet of each subcatchment. The second component – the *routing module* – links the contribution of all subcatchments, and represents the in-stream transport and fate processes of the chemicals.

We assume that subcatchments are laterally disconnected from each other, and therefore simulations of the substance transfer module can be run separately for each subcatchment. Subsequently, the routing module integrates all outputs of the substance transfer module by processing subcatchments from up- to downstream.

For the routing, the main river (and optionally also tributaries) is split into river segments (see Appendix A2A10, Fig. S2S9). Each segment receives input from upstream and lateral directions as well.

3.2 Substance transfer module

This module consists of several independent parts that describe the transfer of chemicals from the different pesticide sources in the catchment. In particular, it consists of the *iWaQa* model describing substance transfer for herbicides (Section 3.2.1) and another for biocides (Section 3.2.2). These models treat subcatchments as spatially lumped units. The models are very parsimonious such that they only require one to eight empirical, yet global model parameters per simulated chemical (Table 1). All other model inputs consist of (generally) available statistical data on chemical consumption, spatial data on land use and hydro-climatic time-series.

3.2.1 Substance transfer for herbicides

This section describes first the system of the herbicide model and subsequently the input and output of the system.

This model consists of two spatially lumped storage terms representing the dissolved and sorbed fractions of the total herbicide mass $M(t)$ [g] in the topsoil layer of agricultural fields in the subcatchment. The first storage is the mass dissolved in the pore water $M_w(t)$ [g] being instantly available for release to the river. The other represents mass adsorbed to the soil matrix $M_s(t)$ [g] and is unavailable for immediate release.

165 The exchange between the two storages is described by two kinetic rate parameters: sorption to the soil matrix is described with the transfer rate k_{w-s} [d^{-1}] and the reverse flux with k_{s-w} [d^{-1}], respectively. Both stocks degrade according to first-order kinetics with decay rate k_{deg} [d^{-1}].

The mass balance and the two first-order differential equation describing the change in stock of herbicide masses $M(t)$ in the system are given by:

170

$$M(t) = M_w(t) + M_s(t) \quad (1)$$

$$\frac{dM_w}{dt} = \rho \cdot \dot{M}_a(t) - k_{w-s} \cdot M_w(t) + k_{s-w} \cdot M_s(t) - k_{deg} \cdot M_w - L_{herbicide} \quad (2)$$

$$\frac{dM_s}{dt} = (1 - \rho) \cdot \dot{M}_a(t) - k_{s-w} \cdot M_s(t) + k_{w-s} \cdot M_w(t) - k_{deg} \cdot M_s \quad (3)$$

175 where $\dot{M}_a(t)$ [$g d^{-1}$] the **rate of mass rate** applied in the catchment during the application period and ρ [-] represent the fraction of the applied mass that is immediately available for transport, **such that it can be directly mobilised when it rains**.

The output $L_{herbicide}$ [$g d^{-1}$] is the herbicide load released from the current application at the outlet of the subcatchment.

Input:

180 Crop development and hence also the timing of herbicide applications is strongly controlled by temperature conditions in any particular year. As application dates are generally unknown, a temperature sum models is used to simulate crop growth and the related herbicide applications, which is linked to specific growth stages of the crops. In particular, we assume that application of herbicides starts when the daily temperature sum at a given location reaches a crop-specific temperature threshold (Honti et al., 2017). Daily mean values of temperature are summed up ($T_{sum}(t)$), though a restart is forced after

185 freezing days. Once the objective temperature T_{obj} is reached, 1/14 of the total application mass is applied on each following rain-free day until the total application mass is depleted. Herewith the selection of a universal application date is avoided and the method accounts for regional climatic differences.

Output:

The concept to describe the transfer of the applied herbicides from the fields to the river is based on the empirical observation that herbicide concentrations increase with flow during discharge events during the application period (Leu et al., 2010). Mechanistically this can be explained by the occurrence of fast transport processes (with high herbicide concentrations) such as surface runoff and fast subsurface flow through drainage systems or macropores (Leu et al., 2004b) during discharge events. Hence the concentration (C [$g m^{-3}$]) in the river is described – in a first approximation – as proportional to the discharge $Q(t)$ [$m^3 d^{-1}$] in the case of a recent application on the fields; the load [$g d^{-1}$] increases

195 quadratically with discharge:

$$C_{herbicide}(t) = \alpha(t) \cdot Q(t) \quad (4)$$

$$L_{herbicide}(t) = C(t) \cdot Q(t) = \alpha(t) \cdot Q(t)^2 \quad (5)$$

200 where α [$g d m^{-6}$] is the proportionality coefficient relating the magnitude of the discharge to the released loads.

The proportionality coefficient depends on $M_w(t)$, the mass dissolved in the pore water and instantly available for release:

Formatted: Font color: Text 1

$$a(t) = \varepsilon \cdot M_w(t) \quad (6)$$

205 where ε [d m^{-6}] becomes a catchment-independent, empirical loss factor that needs calibration for each chemical (see sec. 4.2).

Certain herbicides are present in significant concentration outside of the application period too (see for example (Leu et al., 2004b)). Therefore, we added a constant background concentration (C_{back} [g m^{-3}]) to the substance transfer model. This step
210 was essential to ensure a proper calibration of the model. By doing so we implicitly assume a constant concentration of herbicides independent of the application period, representing e.g. other, not seasonal sources or a general presence in the baseflow due to the long-term persistence of pesticides in groundwater. Thus, the total released load of the system becomes:

$$L_{release}(t) = C_{back}(t) \cdot Q(t) + \varepsilon \cdot M_w(t) \cdot Q(t)^2 \quad (7)$$

215

3.2.2 Substance transfer for biocides

Biocides are applied in the urban settlement on façades, flat roofs, basement seals and underground parking lots. Due to the potential year-round application and the long-term protection purpose of biocides, it is assumed that the stock in the urban settlement is constant over time (Wittmer et al., 2010).

220 The leaching of biocides in urban areas is a complex process and several studies provide quantitative information on loss rates, dynamics and driving factors (Jungnickel et al., 2008; Burkhardt et al., 2008; Wittmer et al., 2011). The process is mainly driven by precipitation when water flows over the treated surfaces and it was observed that concentration patterns of urban compounds follow the rainfall pattern more than the river discharge (Wittmer et al., 2010). Therefore, the current model simplifies the processes by assuming the release being proportional to precipitation and instantaneous transport to the
225 rivers. The following equation thus describes the resulting modelled load $L_{biocide}(t)$ [g d^{-1}] to the rivers:

$$L_{biocide}(t) = M_a \cdot \beta \cdot P(t) \quad (8)$$

230 With M_a [g] the total mass present in the catchment within the model period, β [m^{-1}] the substance-specific loss rate (to be calibrated, see below), $P(t)$ the precipitation [m d^{-1}]. [The assumption of instantaneous transfer to the stream may cause some timing errors if compounds have residence times that are longer than the model time step \(e.g. in wastewater treatment plants\) but see the findings on routing effects in sec. 5.2 .](#)

3.2 Routing module

235 *Load aggregation*

Concentrations of micropollutants at the outlet of any catchment composed of several subcatchments are predicted by aggregating the loads from the output of the substance transfer module and division by the actual total discharge. The approach considers the local availability of sources and the spatial distinctions of the driving factors (discharge or precipitation). However, instantaneous aggregation assumes no in-stream losses, such as degradation, sedimentation or

240 diffusion taking place during the transport. Furthermore, it implies that the temporal resolution should be larger than the longest travel time of a component during a rain or discharge event. Otherwise the concentration dynamics are affected.

A special situation is given by the presence of the large pre-alpine lakes (Lake Constance, Lake Lucerne etc.) in the river network. Because of the long water residence time in these water bodies (months to years), the concentration dynamics in the lake outlet are strongly dampened and differ substantially from other river sections. To account for these different dynamics, we simulated the input into each of these lakes separately by the substance transfer module. We assumed complete mixing into one year of discharge and used the resulting concentration as a constant value in the river water flow out of the respective lakes. The load varied accordingly with discharge from the lake. A different case was Lake Biel, which was not treated as a mixing reactor because of the short spatial distance between the inflow and the outflow of the river Aare.

Routing with AQUASIM

250 In larger river basins the effects of travel time, dispersion and degradation during pollutant transport in the river system become more important. The assumption of instant arrival of pollutants at the outlet within daily time steps does not hold true anymore and hydraulic routing becomes indispensable.

To that end the load output from the substance transfer module was used as input into the program AQUASIM (Reichert, 1994) that was used for describing the transport and fate processes within the main rivers. Flow was described with the kinematic wave approximation of the St. Venant equations. Transformation and sedimentation through sorption was neglected because the model compounds are sufficiently stable and show only weak sorption. ~~(see also Honti et al. (in preparation)).~~

4 Methods

4.1 Model input data

4.1.1 Discharge, precipitation and temperature

Hourly discharge data was obtained for 1033 stations from federal and national agencies (see Appendix A4, Table S4S2) to derive two kinds of discharge time series for all subcatchments. The first, termed local runoff, refers to surface and subsurface runoff originating from the specific subcatchment and is used in the substance transfer module for herbicides. The other is the streamflow at the outlet of a subcatchment required in the routing module to calculate the concentrations of catchments or as input to AQUASIM. For headwater subcatchments without any further upstream connections, the local runoff is identical to the streamflow.

270 Time series of local runoff are derived from the records of gauging stations measuring rivers with a Strahler stream order (Strahler, 1957) less than five (804 out of 931 or 86 % of the available gauging stations). Using gauging stations at larger rivers would not accurately reproduce the high temporal variations of the local runoff. The recorded discharge is allocated to the subcatchments upstream according to the drainage area ratio method, which assumes that discharge scales proportional to catchment area (Hirsch, 1979). Unfortunately, many subcatchments remain ungauged hereby. On one hand this method does not provide time series for subcatchments downstream of the stream gauges with Strahler order larger than four, on the other hand numerous ungauged tributaries join the river network downstream of the selected stream gauges. In both cases a nearby reference stations (with Strahler order < 5) is selected and the area ratio method is applied to calculate local runoff. 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

280 ~~most correlated discharge at the ungauged catchment based on calculating the correlation between stream gauges and choosing the station with the most correlated discharge at the considered catchment.~~

The stream flow time series for all subcatchments were deduced in a similar way. Upstream of stream gauges with Strahler less than five, the discharge is allocated according to the drainage area ratio and accumulated ~~towards~~ downstream. The discharge of any stream gauge is passed on to the downstream subcatchments and accumulated with the streamflow of converging tributaries. Likewise to the local runoff, the streamflow for ungauged tributaries is adapted from reference stations selected with the map-correlation method.

Hourly precipitation data for the study area is available for Switzerland from MeteoSwiss CombiPrecip (Sideris et al., 2014) and for the rest of the Rhine basin from RADOLAN (Bartels et al., 2004), a product of the German Meteorological Office (DWD). Both are raster datasets (with a spatial resolution of 1 km²) computed using geostatistical combination of radar sensing and rain gauge measurements. Small temporal gaps in the precipitation data or uncovered parts in the French region were filled with data from the nearest available rain gauge. Additional data from rain gauges is available for Luxembourg and France. By intersecting the raster cells with the subcatchments, the most accurate conversion was achieved with the area-weighted mean of the overlapping grid cells within a subcatchment.

Raster temperature data with daily mean values are retrieved from MeteoSwiss TabsD (Begert et al., 2003) with a spatial resolution of 0.02 deg (~2.3x1.6km) and from the European dataset termed E-OBS (Haylock et al., 2008) with a coarser resolution of 0.25 deg (~27.8x18.8km). Both datasets are spatial interpolations of monitoring stations.

Given that the Swiss temperature dataset has a finer grid size than the average area of the subcatchments (8.8 km²), it allowed for estimating reliable mean temperatures for all subcatchments in Switzerland. The grid size of the E-OBS temperature data was significantly larger than the average subcatchments. The spatial resolution of the E-OBS temperature data set was therefore refined using a Digital Elevation Model with a grid size of 1 km² (the DEM was obtained from the GMES RDA project, EEA, 2013). In particular, the deviation between the altitude of the DEM cells and the E-OBS cells was calculated. ~~From these altitude deviations, temperature values were corrected based on a temperature decrease by -0.0065°C/m altitude increase. These deviations were multiplied with a temperature lapse rate of -6.5 °C/km~~ and added to the temperature values of the E-OBS cells. Thus a gridded temperature model with a resolution of 1 km² was obtained.

4.1.2 Land use data

Herbicides are applied on specific crops, therefore detailed, spatially distributed agricultural land use data were required. The dataset “Agricultural Landuse2000” from the JRC AFOLU project (Leip et al., 2007) classifies agricultural land use into 30 crops and for a grid with a resolution of 1 km² by combining remote sensing with statistical information of the agricultural production. ~~Because there was no data set available reflecting the most recent situation, we checked whether there have been major shifts in agricultural land use with the spatially lumped data on the 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 mostly little 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 (see Fig. S4). For Switzerland, more recent land use (2004 – 2009) and crop statistics (2010) were available and used.~~

This European dataset on agricultural land use does not cover Switzerland. In order to have a dataset with the same crop categories and a similar spatial resolution, a harmonized dataset was created from the Land Use Statistics of Switzerland

320 (Swiss Federal Statistical Office FSO, 2012) and the census of agricultural enterprises (Swiss Federal Statistical Office FSO, 2011). The cultivation areas of 60 listed crops reported in each municipality in the census were distributed on the grid cells of the Land Use Statistics belonging to the 3 agricultural land use classes, leading to an average fraction of cultivated area of crop l per grid cell in community k :

$$325 \quad W_k^{(l)} = \frac{a_k^{(l)}}{a_k^{(tot)}} \quad (9)$$

$W_k^{(l)}$ [-] is the average fraction of crop l being cultivated in a single grid cell belonging to community k . The $a_k^{(l)}$ [ha] is the cultivation area of crop l (reported in the census) in municipality k , $G_k^{(tot)}$ [ha] is the sum of the area of all agricultural grid cells in community k . The 60 crop categories of the census are merged to the 30 categories from the European “Agricultural Landuse2000”, thus a consistent database is accomplished with a comparable approach of distributing statistically reported areas to spatial land use data.

Land cover of housing and settlements is available with vector based maps, where every building is precisely represented by a polygon and in some cases with knowledge about its height.

- 335
- France: Institute géographique nationale (IGN) BD TOPO® (with height)
 - Germany: Arbeitsgemeinschaft der Vermessungsverwaltungen Deutschland (ADV) ALKIS®
 - Luxembourg: Administration du Cadastre et de la Topographie (ACT) BD-L-TC
 - Switzerland: Federal Office of Topography (swisstopo) swissTLM 3D (with height)

340 Façade surfaces are calculated by multiplying the contours of buildings with their height where available (CH, FR). For the other countries (DE, LU) the façade areas within a subcatchment are estimated from the footprints areas and the population. Footprint and façade follow a linear relation, whereas the relationship between population N_{pop} [-] and façade A_{fac} [m²] appear to be polynomial. With the Swiss data the following regression was obtained:

$$345 \quad A_{fac} = 1.55 \cdot A_{foot} + 1.45 \cdot 10^5 \cdot N_{pop} + 6.20 \cdot 10^{-4} \cdot (N_{pop})^{0.49} \quad (10)$$

This regression was validated with the French data achieving reasonable results and finally used to calculate the façade areas in Germany and Luxembourg (see Supplementary Information, Appendix A6, Fig. S5).

4.1.3 Model compounds, use and sale data

350 Five model compounds (see Tab. S1) have been selected for this study: three agricultural herbicides (isoproturon (IPU), S-metolachlor (MEC), terbuthylazine (TBA)) and two (dual use) biocide (carbendazim (CBZ), diuron (DIU)). The biocides are mainly used in urban environments to protect materials. They may also have some agricultural use in some regions of the basin (e.g., in Switzerland) but the usage is of minor relevance and is neglected here.

355 Use and consumption data for the chemicals are not available in a spatially distributed manner. To provide input for all spatial model units, we proceeded in two steps. First, we obtained statistical data on use/consumption data for regions or countries. Subsequently, we downscaled these statistical data based on land use or population.

Annual sales data of herbicides were available from the countries Switzerland (Agroscope ZA-AUI, (Spycher and Daniel, 2013)), Germany (Federal Office of Consumer Protection and Food Safety, (Federal Office of Consumer Protection & Food

360 Safety BVL 2008 - 2012)) and the French regions Alsace (Office national de l'eau et des milieux aquatique, (Office national
 de l'eau et des milieux aquatique ONEMA, 2014)) and Lorraine (Groupe Régional d'Action contre la Pollution
 Phytosanitaires des Eaux Lorraine, (Groupe Régional d'Action contre la Pollution Phytosanitaires des Eaux Lorraine
 GRAPPE Lorraine, 2005)) for the years 2008-2012 (except the study for Lorraine was only issued for 2005). The spatial
 coverage area of the statistics varied strongly ranging from 357'300 km² for Germany to 8'330 km² for Alsace. The Swiss
 365 dataset only provided coarse ranges of substance sold per year from which the mean values were used.
 Only one source for the use and sale of biocides was at hand. The survey of Burkhardt & Dietschwiler (2013) investigated
 the consumption rates in Switzerland of various biocides in antifouling paints, masonry and wood protection agents. The use
 rates have been applied to the entire study area.

370 The mass distributed on the agricultural fields respectively applied on houses of each catchment was estimated by
 downscaling regional or national sales data \dot{M}_{tot} [g d⁻¹] with the ratio of the local application area (area within a
 subcatchment) A_a [ha] to the total application area A_{tot} [ha] (total area within the considered sales study):

$$\dot{M}_a = \frac{A_a}{A_{tot}} \dot{M}_{tot} \quad (11)$$

375 The application area was distinct for use classes and substances. For herbicides it was the sum of possibly treated agricultural
 land use areas, more specifically the crops for which a substance is authorized and primarily used. [The resulting spatial
 distribution of estimated input is depicted in Appendix A6, Fig. S5.](#) Biocides were applied on façades of a building. The sum
 of the respective building surface composes the application area. [Because of the lack of spatially distributed biocide use data,
 380 the spatial distributions of CBZ and DIU are identical \(see Appendix A6, Fig. S6\).](#)

4.2 Calibration of the catchment model

4.2.1 Calibration sites

To calibrate the model, data from field studies was used that provided simultaneously data on application amounts of
 substances as well as on losses to the rivers. Such studies are rare and we used the following studies situated in the
 385 northeastern part of Switzerland. The sampling campaigns from Gomides Freitas et al. (2008) and Doppler et al. (2012)
 measured herbicide concentrations at the small-scale agricultural catchments Summerau and Ossingen, respectively, after a
 controlled herbicide application. Wittmer et al. (2010) monitored the mass and dates of herbicide applications in a slightly
 larger catchment Mönchaltorf (25 km²) with mixed land use. The biocide application was estimated with product and
 statistical information. Subsequently the losses from the catchments were measured at the outlet of the catchment.

390

4.2.2 Calibration procedure

The substance-specific parameter sets for herbicides $\theta_{herbicide} = \{\varepsilon, C_{back}, \rho, k_{w-s}, k_{s-w}, k_{deg}, T_{obj}\}$ and for biocides
 $\theta_{biocide} = \{\beta\}$ cannot be measured and require calibration. Parameter T_{obj} regulating the timing of herbicide application was
 only calibrated in the case of Mönchaltorf where regular application occurred at the farmers' chosen timing. At Ossingen and
 395 Summerau the application was experimentally controlled and therefore a calibration of T_{obj} would be meaningless.

The model parameters were calibrated using a Bayesian inference approach. The likelihood function accounted for the
 parameter uncertainty and the structural model errors. For herbicides, model errors were assumed to deviate stronger during
 the application season. Therefore an error-scaling function was added depending on the substance input to the system and a

400 driver imitating the approximate substance application to the fields. 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. The additional parameters to calibrate resulting from the error-scaling function were $\theta_{herbicide,error} = \{\mu, \sigma_{error}\}$ where μ is a scaling factor for the substance input and σ_{error} the calibrated standard deviation of the total model error. For the biocides the error variance was assumed to have no seasonality.

405 Measured peak concentrations of herbicides in the calibration studies occurring before the monitored application period were excluded from the calibration procedure as they represent accidental spills or runoff from hard surfaces. As such events are not represented in the model, including them would have spoiled the identification of model parameters.

The likelihood function used in this study is based on the assumption that Box-Cox transformed (Box and Cox, 1964) time series of concentration data C lead to independent and identically distributed normal errors as described in Honti et al., (2017). The corresponding likelihood function is as follows:

Formatted: Font color: Text 1

Formatted: Font color: Text 1

$$p(C_{obs}|\theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N \left(g(C_{obs,i}) - g(C_{mod,i})\right)^2\right) \prod_{i=1}^N \frac{dg(C_{obs,i})}{dC} \quad (13)$$

415 where σ^2 is the error variance, N is the total number of observations in all subcatchments, C_{obs} and C_{mod} are the observed and the modelled concentrations for the data point i . The transformation $g(\cdot)$ is the Box-Cox transformation used to remove the heteroscedasticity of the residuals:

$$g(C) = \frac{C^{\lambda-1}}{\lambda} \quad (14)$$

420 The parameter λ was set to 0.3.

The Jacobian of the transformation $\frac{dg(C_{obs})}{dC} = \prod_{i=1}^n C_{obs,i}^{(\lambda-1)}$ was required to compensate the distortion of the likelihood by using the transformed variables.

4.2.3 Prior distributions

425 Priors for the substance-specific loss rates were estimated based on reported information in the calibration studies (see Appendix A8, Table S4). Estimation for the substance-specific ε of the herbicide model is based on the reported loss rates from these studies. Neglecting background concentrations the time-averaged concentration \bar{C} during the main loss period from t_0 to t_{0end} is given according to Eq. 4 as

$$430 \overline{C_{herbicide}} = \frac{\varepsilon}{(t_{end}-t_0)} \times \int_{t_0}^{t_{end}} M_w(t) \times Q(t) dt: \quad (15)$$

Based on measurements, \bar{C} can also be expressed as:

$$\overline{C_{herbicide}} = \frac{M_a \times \Delta\tau \times Ir_{study}}{\int_{t_0}^{t_{end}} Q(t) dt} \quad (16)$$

435

where \bar{M}_a is the average application rate in the catchment, $\Delta\tau$ is the duration of the application period, lr_{study} is the empirically observed loss rate from the study. From Eq. 14 and 15, it follows that ε can be approximating as:

$$\varepsilon = \frac{\bar{M}_a \times \Delta\tau \times lr_{study}}{\int_{t_0}^{t_{end}} Q(t) dt \times \int_{t_0}^{t_{end}} M_w(t) \times Q(t) dt} \approx \frac{\bar{M}_a \times \Delta\tau \times lr_{study}}{(t_{end}-t_0) \times \bar{M}_w \times (\bar{Q})^2} \quad (17)$$

440

where \bar{Q} is the mean discharge during this period, and \bar{M}_w is the mean [mass available for transport](#) calculated using the known application pattern and a first-order approximation for the sorption and decay.

Priors for the substance-specific loss rates of the biocide model was the total loss rate reported in Wittmer et al. (2010) divided by the yearly sum of precipitation. Having multiple study catchments or ranges of loss rates allowed to calculate a
445 distribution of the priors for ε and β ([see Appendix A9, Table S4 and S5](#))-

Prior distributions for the parameters describing pesticide fate in the soil (ρ , k_a , k_{w-s} , k_{s-w} , k_{deg}) were derived from field experiments. The equations are fitted to the Freundlich adsorption isotherms with time-varying sorption coefficients measured in soil samples ([Gomides Freitas et al. 2008](#)).

The maximum of the posterior parameter distribution was found by performing a Nelder-Mead simplex optimization. The
450 maximum likelihood parameter set was used as a prior for the Markov chain Monte Carlo (MCMC) simulation using the Metropolis algorithm (Gelmanman, 1997). The developed posterior parameter distributions were used to predict the parameter and model uncertainty. The procedure was repeated for every calibration site separately.

4.3 Model validation and routing

Several comprehensive sampling campaigns from the Swiss “National Surface Water Quality Monitoring Program ~~—~~
455 [Special Campaigns — NAWA SPEZ](#)” (Federal Office for the Environment FOEN, 2013) and data from a continuous monitoring station were selected to evaluate the model.

The first campaign (~~called NAWA SPEZ 2012~~) comprised five catchments (Fig. 2) ranging from 39 to 105 km² ~~and~~
~~have with~~ varying extents of urban and agricultural influences (Appendix A7, Table S3). The measurement campaign was accomplished from March to July 2012 with biweekly time-proportional mixed samples (Moschet et al., 2014).

460 The second survey was the “~~National long-term surveillance of Swiss rivers~~[National River Monitoring And Survey](#)” termed NADUF, where weekly or biweekly mixed-samples were taken during 2009 (Stamm et al., 2012). The monitoring sites were in the north-eastern part of Switzerland and quantified the concentrations of several organic micropollutants in five nested catchments. These nested catchments have a large range of size from 74 km² to 14'718 km² comprising between 22 and 2554 subcatchments ([Fig. S8](#)).

465 A third validation was conducted with data for 2011 from the continuous measurement program of the International Rhine Monitoring Station (IRMS) near Basel. With five probes distributed over the cross-section, daily discharge-proportional pollutant levels are evaluated. The upstream area of the Rhine at this point covers almost 36'000 km² including the sub-basins Alpine Rhine, Lake Constance, High Rhine and Aare.

470 Modelled hourly concentrations were adapted to the sampling periods of the respective validation surveys. According to the aggregation periods of mixed samples in the measurement surveys, the modelled concentrations were averaged over the sampling time periods, such that the resulting time series were fully comparable.

The issue of routing arises for larger catchments where the transport time is longer and also the processes along the way
475 become more significant. For the sites of the NADUF survey the concentrations at the outlets were first modelled with load

aggregation and in a second step river segments were defined where the routing with AQUASIM was calculated. Thus the influence of a physically-based hydraulic routing can be compared to the situation where in-stream transport and processes are neglected.

In the case of the IRMS, measuring a large sub-basin of the Rhine, the catchment model is applied for 5950 subcatchments.

480 | Downstream of the lakes the substance transport was modelled with AQUASIM for the larger rivers (such as Rhine, Aare; Appendix 9 Appendix 10, Fig. S6S9). The simplistic approach with load aggregation was applied on this large scale as well.

4.4 Model predictions within the Rhine basin

The calibrated model was finally applied to the Rhine and the major tributaries to characterize the pollutant dynamics of herbicides. These simulations were real predictions without any further adjustments of model parameters. Due to the lack of
485 | statistical input data of the use of biocides in France and Germany predictions for the Rhine basin were not possible for carbendazim and diuron.

4.5 Technical implementation

The *iWaQa* model is written in C++ and the outputs are time series of concentrations, parameter estimations, posterior parameter distribution from the MCMC or matrices with the concentration predictions with the posterior parameters. Within
490 | a Python framework, i) the input for the substance transfer module is generated, ii) the substance transfer module runs the *iWaQa* model for the entire Rhine basin and iii) the two routing options are executed (see Appendix A1). Data preparation and analysis is ~~effectuated~~ performed with the programming language R (R Core Team, 2017).

All modules are executable individually. Preprocessing succeeds within 30 minutes to sort the hourly input data for all
495 | 18'240 subcatchments of the Rhine basin on an Intel x86 8-core processor. The substance transfer module takes approximately an hour to run and sort the output by both, subcatchments and time steps. Run times of the routing options differ substantially depending on the size of the considered catchment and the parameterization of AQUASIM. Generally the load aggregation is calculated within a few minutes and the simulation of the main tributaries of the Rhine basin with AQUASIM is completed within 6 hours.

4.6 Model evaluation

500 | Besides the likelihood used for parameter calibration, there are many metrics for evaluating model performance of hydrological and water quality models (Jachner et al., 2007; Smith and Rose, 1995; Reusser et al., 2009; Moriasi et al., 2007).

Out of those, we have selected some frequently used statistics (Table 3 Table 3) that allow for a comparison with other studies. In addition, we have included some metrics that are more specifically designed to analyse aspects, which are of special relevance for this work. These measures include the Geometric Reliability Index of the cumulative distribution of the
505 | simulated concentrations to see how well the overall concentration level is met or the fold difference between the observed and simulated maximum concentration during the simulation period (see Table 3).

5 Results

5.1 Calibration

The calibration was carried out for all catchment-compound combinations for which observations are available (see Table 2).

510 | For the agricultural herbicides this provides several alternative calibration sets (Tab. S6, S7). -The final decision of which set to use for further predictions was based on the performance in the validation step with the NAWA SPEZ sites (see below).

For the agricultural herbicides, the calibration resulted in a reasonable simulation of the observed concentration dynamics (Figure 3, Supplementary Material Fig. S7, S8, S12, S33 - S34, Tables S8 - S10). The calibrated uncertainty bands also followed the expected seasonal patterns: they were large during the application periods and decreased with time. The model, however, poorly captured the exact timing of the concentrations as one can see from the low Nash-Sutcliffe (NSE) coefficients (ranging between -0.05 and 0.62, median = 0.38; see Appendix A13, Table S10). Despite these deviations, the correlations between observations and simulations were reasonable (range between 0.30 and 0.85, median = 0.68).

For the biocides, the model predicted a rather uniform distribution of concentration peaks around the year reflecting the precipitation patterns. The observations however, suggest a bi-modal seasonal pattern with higher concentrations in spring and fall. This pattern resulted in low correlations (r of 0.30 and 0.37; see Appendix A13, Table S10) and poor NSE values (-0.05 and 0.08), which is also reflected by the poor relationship between the P/Q ratio and the observed biocide concentration over the year (Fig. S10). A possible reason for this temporal pattern is a seasonal biocide application. However, there is no data available for testing this hypothesis.

The residuals pointed to systematic deviations between observed and modelled concentrations (Figure 4, Fig. S11). The data group into two clusters. One of the clusters showed systematic underestimations of the observations, while the other showed the opposite. Comparison with the time-series revealed on the one hand that for most compounds, the highest observed concentrations peaks were (substantially) underestimated during calibration (see for example S-metolachlor or terbuthylazin in Figure 3). These peak concentrations were underestimated by 13% to 83% (Table 4, Table 4). On the other hand, the second cluster of data points indicates that concentrations of some (smaller) events were overestimated. This pattern suggests that the model structure did not capture the full dynamic range of the pesticide concentrations.

Despite these limitations, the concentrations were reasonably well represented by the model. The Geometric Reliability Index GRI indicates that the predicted concentrations of the agricultural herbicides were within a range of 1.9 to 2.5 of the observations (Figure 5). When being interested in how well the cumulative concentration distributions are simulated (ignoring the timing) – this generally relevant for water quality assessment - these values range between 1.4 and 2.2. As can be seen from Figure 5, the performance for the biocides was considerably poorer but the cumulative distribution was also reproduced better than the concentration time series.

Based on the relative RMSE one can compare the calibration performance across sites. Mönchaldorf and Summerau yielded better calibrations for S-metolachlor and terbuthylazin than the Ossingen data set (Appendix A13, Table S8). The opposite was true for isoproturon. In the case of Ossingen, a long dry period followed after the isoproturon application resulting in very low concentrations without a pronounced peak related to the recent application. This last aspect points to the fact that single calibration data sets may represent special situations hampering the predictive power during normal conditions. The application of S-metolachlor and terbuthylazine in Ossingen for example, took place later – just before an intensive precipitation event. Through direct shortcuts, such as manholes of drainage systems and storm drains, the transfer to the river was accelerated and very high concentrations have been measured (Doppler et al., 2012).

So far, we have compared the observations to the deterministic model predictions. Comparing the observations to the simulations including the prediction uncertainties due to the estimated parameter uncertainty (of the deterministic model) and the total predictive uncertainty accounting for input and model structure deficits reveals that the parameter uncertainty contributes only a small fraction. Taking into consideration all sources of uncertainty leads to uncertainty bands that include most of the observations as can be seen from the cumulative concentration distributions depicted in Appendix A12, A14, Figures S23, S27 - S28, S24.

All calibrated parameters of the deterministic model had priors based on physical reasoning or empirical data, hence the maximum likelihood values are not expected to deviate strongly. This held true for the decay rate, the loss rates (ϵ and β), the background concentration and the objective temperature. The parameters describing the herbicide (de-)sorption processes (initial availability ρ , transfer rates k_{s-w} and k_{w-s}) changed considerably. In general, the sorption coefficient values were higher and degradation rates smaller than ~~in~~^{the} priori estimates, meaning that the available mass for release was smaller but more persistent. For sorption, this could be explained to some degree by different soil-water ratios of undisturbed soils and the conditions during the experimental procedure from which the priors were derived (Gomides Freitas, 2005). However, stronger sorption in the model could also compensate for pesticide applications that were missed by the model.

5.2 Validation

The different calibrated parameter sets were used to predict the corresponding concentrations for the validations case studies. To that end, the model output having a daily resolution was aggregated to the time periods of the real sampling strategies at the respective sites. In contrast to the calibration procedure, the validation step included also the simulation of the compound input. This included the estimation of the applied amounts and the timing of the applications.

For the agricultural herbicides, several calibration data sets were available. All of them were first tested on the NAWA SPEZ sites. Based on their performance, one set per compound was used for simulating the larger NADUF and IRMS sites. Based on the correlation coefficients and the NSE criterion the parameter sets calibrated at Mönchaltorf for the compounds isoproturon and terbuthylazine and the parameter set from Summerau for S-metolachlor were used for the other catchments (see Appendix A13A14, Table S8).

At the IRMS, the validation of the model was partially restricted due to the low concentrations that often remained below the limit of quantification (LOQ) of 5 ng/l for S-metolachlor and terbuthylazine. Nevertheless, concentrations were high enough to evaluate the model performance during the application period.

The quality of the predictions varied between compound use class and the validation catchments. The GRI values demonstrate that the agricultural herbicides were simulated better than were the biocides (Figure 5, compare respective red and blue distributions). The (cumulative) distribution of observed concentrations was better represented by the model as compared to the actual time series. Interestingly, the model performed better in the larger catchments (Figure 5) despite the fact that calibrations were up-scaled to areas that are between four and 70'000 time larger than the calibration catchment (Table 2). This might be explained by averaging out regional differences and variabilities in local application dates – hence also input uncertainty - across larger scales.

The quality of the predicted maximum concentrations changed from the calibration to the validation step. While the values were systematically underestimated during calibration, this pattern changed substantially for the validation. Depending on the site-compound combination, the maximum concentrations were either clearly under- or overestimated (Table 4). Irrespective of the sign of the deviation, the fold difference between observed and simulated concentrations – indicating by which factor observations were over- or underestimated – mostly ranged between one and four (Figure 6). However, there were a few cases of extreme deviations because of either the observation of a pronounced and very high peak or very low measured values hardly exceeding the observed background. Again, the model performed better for the herbicides where for 50% of the predictions (site-compound combinations) the maximum concentrations were predicted within a factor of 2.0 deviation from the observations. For the biocides, the value was larger (> 3.0). We observed also clear compound-specific patterns such as systematic over-estimation of diuron concentrations (see e.g., Appendix A12A13, Figure S28S32).

As during the calibration step, the Nash-Sutcliffe values were low pointing again to the problem of properly simulating the exact timing of concentration peaks (Figure 7S33). This was very pronounced for the biocides. The correlation coefficients provided a mixture picture. For some compounds such as diuron, the correlations coefficient range between 0.29 and 0.68 (median = 0.56) for the NAWA SPEZ and NADUF sites. For others such as carbendazim or isoproturon the correlation was very variable especially between the NAWA SPEZ sites (see Appendix A13A14, Table S10). At the station on the Rhine in Basel, the correlations varied between being non-existing to fairly strong (isoproturon: $r = 0.84$ assuming load aggregation across the Rhine basin).

605 *Effects of routing*

For the IRMS measuring site, we compared the performance of the simple load aggregation procedure and the explicit routing with AQUASIM (see Table S10). Differences between both approaches were moderate. The routing yielded better results because some of the pronounced concentrations peaks predicted by load aggregation were substantially smoothed. Therefore, the maximum concentrations were overestimated to a lesser degree. The median difference between observed and simulated maximum concentrations with and without routing were 3.1 and 3.4-fold, respectively (averages: 2.6 and 4.8, respectively). The slightly better NSE values also suggest a better performance with an explicit routing. These results provided evidence that at the scale of such large basins of 30'000 km² and beyond the explicit routing makes a relevant difference for pesticides studied at a daily resolution.

5.3 Predictions for the Rhine basin

615 Based on the findings reported above on the effects on routing, we only report the findings based on AQUASIM for the predictions of the main tributaries (Aare, Neckar, Main and Moselle) and the further measuring sites downstream of Basel. The total river length for which the routing was explicitly simulated with this module was 1773 km. We focus here on the three herbicides (isoproturon, S-metolachlor and terbuthylazine) because for them a minimum set of observations was available.

620 The observed isoproturon concentrations revealed the two peaks in spring and fall as measured also at the other locations (Figure 8Figure 6). The model predicted the timing of the spring peak very well. Also the absolute concentrations level of the peak was simulated well (within 30% of the observation). Concentrations during the summer months were slightly underestimated; the fall peak was missed because no application was included in the model (see above).

The comparison of the simulated chemographs along the Rhine show some slight temporal shifts of the peaks caused by different application periods. Despite of the size of the basin however, these shifts due to varying phenology were small corresponding to a few days only.

630 The simulations show very similar patterns for the other two herbicides in the different tributaries (see Appendix A11A13, Figure S22S25 - S26). The time shifts between the different sub-basins were also very small. Unfortunately, these findings cannot be tested against observations because the LOQ (10 ng/L and 50 ng/L for S-metolachlor and terbuthylazine, respectively) were too high. Nevertheless, the observed peak concentration for S-metolachlor at Lobith (20 ng/L) was close to the simulated value of 15 ng/L. For terbuthylazine, all simulated values at Lobith remained below the LOQ. This demonstrates at least that the concentrations were not overestimated. This contrasts with the results at Basel where the model predicted a maximum concentration 1.9 times the actual observation.

In our simulations, we have assumed that the compounds behave like conservative tracers without degradation or sorption taking place. Although this is not completely true, the travel times through the river network is so short that relevant dissipation can be expected to be negligible for the model compounds considered in this paper (Honti et al., in preparation).

6 Discussion

6.1 Model performance

We presented here a series of predictions for herbicide and biocide concentrations in streams without any local calibration or model adaptations. In this sense, the results correspond to predictions in ungauged catchments covering tens of thousands of km² based on calibration catchments covering less than 30 km² in total. Despite the challenges that go with this task, the model validation demonstrated that the concentration levels could be predicted within a factor of two to four for 50 to 75% of the predictions. This is comparable to what has been observed for models predicting concentrations of micropollutants from point sources (Johnson et al., 2008). The seasonality of the herbicide concentration peaks was well represented while that of biocides was not well reflected (see below). Overall, the results suggest that such a parsimonious model can be used as a meaningful screening tool to identify potential hotspots in river networks. 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. Models of a similar degree of parsimony have been developed for point source pollution (e.g., Ort et al., 2009) but are largely lacking for compounds with rain-driven input dynamics.

Despite these positive aspects, one has to be clear about the limitations of the model and the resulting predictions. Deficiencies are obvious when evaluating the performance metrics. The NSE or correlation coefficients obtained are low compared to values typically called satisfactory or good for hydrological models. However, our results need to be put into the context of comparable water quality studies dealing with diffuse pollution. As pointed out by Pullan et al. (2016) there is a lack of studies in this field reporting quantitative performance metrics such as NSE or r values. However, studies that do report such values demonstrate that the low NSE or correlation values of our work are in similar ranges of what others have described. Table 5 and Fig. S34 summarise a selection of such findings from a number of model applications (e.g., SWAT, INC-P and others), which are much less parsimonious than the *iWaQa* model used in this study. This comparison indicates that model performance of water quality models achieved so far is generally considerably lower compared to what purely hydrological models can accomplished.

The fact that a parsimonious model such as the *iWaQa* model presented here was able to yield meaningful predictions suggests that the model concept represents the effects of the major drivers controlling the degree and dynamic of biocide and herbicide – and to some degree biocide – inputs into streams. It also indicates that these drivers remain constant over considerable spatial areas and that one can use findings from small study areas to extrapolate to larger basins as long as the first order controls do not strongly change. For the *iWaQa* model as implemented here this means that the herbicide input for example is mainly triggered by discharge events. However, in drier regions it may be possible that point sources play a dominant role (Müller et al., 2002). In this case, the model concept had to be complemented to account for this input pathway as discussed in Honti et al. (2017).

The observation that findings from small catchments can be extrapolated to larger areas in a meaningful manner may be considered a contradiction to earlier work where important spatial differences between herbicide loss rates within catchments were demonstrated (Doppler et al., 2014; Leu et al., 2010). However, the data suggest that spatial heterogeneity at small scales is averaged out at larger ones such that it does not dominate the large scale patterns.

6.2 Model limitations

Despite the positive aspects mentioned above, there are several (major) model limitations one has to be aware of. First, the parsimonious and empirical structure of the model requires compound-specific calibration. This generally implies that field data is available at the catchment scale with sufficiently well quantified input and output fluxes.

While this calibration step is also necessary for other (more complex) models there are also model limitations related to the model structure. During calibration, we have noticed that the model was not able to fully represent the observed concentration peaks (see Table 4). This suggests that the model structure misses important processes that control concentrations during rainfall events. A possible candidate for such a process is drift deposition on roads and subsequent runoff during rainfall (e.g., Lefrancq et al., 2013). Interestingly, this systematic problem during the calibration phase was only partially observed during validation. Possibly this was due to the (much) larger scale of the validation catchments that average over many temporally independent application events.

Other structural model limitations are too high herbicide background concentrations in some sub-basins, the lack of an isoproturon application in fall, or seasonal biocide concentration peaks that were not represented by the model ~~or the lack of an isoproturon application in fall~~. These limitations were rather easy to identify but not very easy to solve. The first problem of herbicide background concentration levels would require a more explicit modelling of the long-term fate of these compounds in the coupled unsaturated and saturated zone. To keep such a model parsimonious one had to test whether these background concentrations could be empirically linked to some simple catchment characteristics. Second, the herbicide application in fall ~~for example~~ is much more difficult to predict compared to the spring application because it not only depends on a single variable such as the temperature sum over the year but it is also influenced by all the climatic variables determining the time of cropping of the previous crop and potential intercropping. In the future, this deficit may be overcome by deriving a stochastic application model based on application data obtained from national surveys. Regarding the third aspect, the ~~For~~ seasonal biocide patterns, we lack any information about biocide use on buildings that could explain the observed seasonality. Targeted surveys on actual use across the year might be a solution. Better input data could then allow to study further structural deficits of this very simple biocide model in more detail.

These examples demonstrate that the model limitations are often a mixture between too simplistic model structure and lack of input data. This agrees with the findings from the error models (see Appendix 14, Fig. S27 – S28). The predictive uncertainty due to poorly identified parameters only explain a small fraction of the deviations between observations and the deterministic model predictions in the calibration phase. The estimated uncertainty for the full error model however, covers most of the data. However, one should not overstate this observation because the fraction of uncertainty assigned to different sources depends on how the error model was formulated (Honti et al., 2014).

One could conclude that a more complex model was warranted to overcome such limitations. While this would be definitely worth considering one should be aware of the severe limitations that come with the input uncertainty regarding the chemicals to be modelled. To illustrate this point, we have quantified the spatial and temporal density of input data needed for the model (Figure 10 ~~Figure 7~~). Compared to the drivers of the hydrological part such as precipitation the density of data on biocide and herbicide input into the system was orders of magnitude lower. While there is hourly precipitation data available on a 1 x 1 km² grid for the entire model domain we could only approximate the herbicide input based on average national sales data. For biocides, one had to rely on a single rough estimate per compound for the entire basin.

Given this level of input uncertainty, it comes as no surprise that the observed concentrations may be substantially over- or underestimated in a given subcatchment. The degree of mismatch between observations and simulations was still in a range that allowed to use the model as a screening tool for identifying potentially critical catchments in a basin. This was probably thanks to the widespread use of the selected model compounds. For less frequently used compounds, one can assume that the input estimates based on sales statistics would be even more uncertain due to e.g. region-specific application patterns.

Accordingly, the predictive uncertainty would increase further.

7 Conclusions

Our findings suggest that even a very parsimonious model with a maximum of eight global parameters that need to be calibrated is sufficient to capture the key drivers and processes for diffuse agricultural herbicide and urban biocide losses reasonably well such as to predict level of peak concentrations within a factor of 2 to 4. This demonstrates that land use as a proxy for compound use, weather data for the timing of herbicide applications and discharge or precipitation as drivers for fast transport are first order controls for diffuse pollution for the compounds in our study area. The results further demonstrate that impact of these factors can be scaled spatially across at least four orders of magnitude (from $< 3 \text{ km}^2$ to $> 30'000 \text{ km}^2$).

At the same time the results also point to clear model deficiencies such as the simulation of background concentrations or the lack of the fall application of certain herbicides. Unfortunately, the analysis of model performance is limited by the lack of adequate validation data that have to combine reliable information on timing and amounts of the use of the chemicals and on concentrations in the streams. Progress in modelling and in measuring will remain closely coupled in this area and mutually benefit from each other.

Finally, it should be recognized that despite using a very parsimonious model, collecting the necessary input data and bringing it into a consistent form across a large water basin such as the Rhine is very time consuming. Hence, sharing model codes and even more importantly the required data will benefit the scientific community by not having to re-invent the wheel.

Data and code availability

The source code and the input data for the models has been placed to GitHub at <https://github.com/moserand/crosswater>. The Supplement related to this article is available online at doi:10.5281/zenodo.556143.

Author contributions. CS designed the initial study design. AM, RS, FF, and CS continuously discussed and guided the project progress. DW, RS, and AM prepared the input data. DW and AM did most of the model coding with essential support from MH. AM and CS did most of the data analysis, figures where provided by RS, AM, and CS. AM and CS prepared the manuscript with contributions from all co-authors.

Competing interests. FF and CS are both on the editorial board of HESS. All other authors declare that they have no conflict of interest.

Acknowledgements.

The CrossWater project was financed by the Swiss National Science Foundation (Grant no. 406140-125866) and builds on previous work funded by the Swiss Federal Office for the Environment (contribution D. Wemyss). Hans-Peter Bader supported the project by regularly discussing important issues during the entire duration of the project. We like to thank Mike Müller for getting us started in Python.

References

- Archfield, S. A., and Vogel, R. M.: Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments, *Water Resour. Res.*, 46, W10513, 2010.
- 755 Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R., Haney, E. B., and Neitsch, S. L.: Soil and Water Assessment Tool. Input/Output File Documentation. Version 2009, 2011.
- Bannwarth, M. A., Sangchan, W., Hugenschmidt, C., Lamers, M., Ingwersen, J., Ziegler, A. D., and Streck, T.: Pesticide transport simulation in a tropical catchment by SWAT, *Environmental Pollution*, 191, 70-79, <http://doi.org/10.1016/j.envpol.2014.04.011>, 2014.
- 760 Bartels, H., Weigl, E., Reich, T., Lang, P., Wagner, A., Kohler, O., and Gerlach, N.: Projekt RADOLAN, Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer), Abschlussbericht (in German), Deutscher Wetterdienst, 2004.
- Beck, M.: Water quality modeling: A review of the analysis of uncertainty, *Water Resour. Res.*, 23, 1393 - 1442, 1987.
- Begert, M., Seiz, G., Schlegel, T., Musa, M., Baudraz, G., and Moesch, M.: Homogenisierung von Klimamessreihen der Schweiz und Bestimmung der Normwerte 1961--1990, Schlussbericht des Projektes NORM90, 170, 2003.
- 765 Behrendt, H., Kornmilch, M., Opitz, D., Schmoll, O., and Scholz, G.: Estimation of the nutrient inputs into river systems - experiences from German rivers, *Journal of Material Cycles and Waste Management*, 3, 107-117, 10.1007/s10113-002-0042-3, 2002.
- Berlekamp, J., Lautenbach, S., Graf, N., Reimer, S., and Matthies, M.: Integration of MONERIS and GREAT-ER in the decision support system for the German Elbe river basin, *Environ. Model. Softw.*, 22, 239-247, 2007.
- 770 Bernhardt, E. S., Rosi, E. J., and Gessner, M. O.: Synthetic chemicals as agents of global change, *Frontiers in Ecology and the Environment*, 15, 84-90, 10.1002/fee.1450, 2017.
- Beven, K., and Kirkby, M.: A physically based, variable contributing area model of basin hydrology, *Hydrological Sciences Journal*, 24, 43 - 69, 1979.
- 775 Borah, D. K., and Bera, M.: Watershed-scale hydrologic and non-point-source pollution models: review of applications, *Transactions of the ASAE*, 47, 789 - 803, 2004.
- Bossel, H.: Understanding dynamic systems: shifting the focus from data to structure, in: *Informatik für den Umweltschutz*, edited by: Hilty, L. M., Jaeschke, A., Page, B., and Schwabl, A., Metropolis Verlag, Marburg, 63-75, 1994.
- 780 Boulange, J., Watanabe, H., Inao, K., Iwafune, T., Zhang, M., Luo, Y., and Arnold, J.: Development and validation of a basin scale model PCPF-1@SWAT for simulating fate and transport of rice pesticides, *J. Hydrol.*, 517, 146-156, <http://dx.doi.org/10.1016/j.jhydrol.2014.05.013>, 2014.
- Box, G. E. P., and Cox, D. R.: An analysis of transformations, *Journal of the Royal Statistical Society, Series B.*, 26 211-252, 1964.
- 785 Brun, R., Reichert, P., and Künsch, H. R.: Practical identifiability of large environmental simulation models, *Water Resour. Res.*, 37, 1015-1030, 2000WR90035, 2001.
- Burkhardt, M., Zuleeg, S., Vonbank, R., Simmler, H., Lamani, X., Bester, K., and Boller, M.: Biocides in facades runoff and storm water of urban areas, Edinburgh, Scotland, UK, 7 p., 2008.
- Burkhardt, M., and Dietschwiler, C.: Mengenabschätzung von Bioziden in Schutzmitteln in der Schweiz--Bautenfarben undputze (PA 7), Holz (PA 8), Mauerwerk (PA 10) und Antifouling (PA 21), Hochschule für Technik Rapperswil, 2013.
- 790 Doppler, T., Camenzuli, L., Hirzel, G., Krauss, M., Lück, A., and Stamm, C.: The spatial variability of herbicide mobilization and transport: a field experiment at catchment scale, *Hydrological and Earth System Sciences*, 16, 1947 - 1967, 2012.
- Doppler, T., Lück, A., Camenzuli, L., Krauss, M., and Stamm, C.: Critical source areas for herbicides can change location depending on rain events, *Agriculture, Ecosystems & Environment*, 192, 85 - 94, 2014.
- 795 Federal Office for the Environment FOEN: NAWA – Nationale Beobachtung Oberflächengewässerqualität. Konzept Fliessgewässer, Federal Office for the Environment FOEN, Bern, 72 p., 2013.
- Federal Office of Consumer Protection & Food Safety BVL Domestic sales and export of plant protection products, in, 2008 - 2012.
- 800 Gamerman, D.: Sampling from the posterior distribution in generalized linear mixed models, *Statistics and Computing*, 7, 57-68, 1997.
- 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:5210.5194/hess-5217-5213-2013, 2013.
- 805 Gomides Freitas, L.: Herbicide losses to surface waters in a small agricultural catchment, Swiss Federal Institute of Technology, Zürich, 143 pp. pp., 2005.
- Gomides Freitas, L., Singer, H., Müller, S. R., Schwarzenbach, R., and Stamm, C.: Source area effects on herbicide losses to surface waters - A case study in the Swiss Plateau, *Agriculture, Ecosystems & Environment*, 128, 177 - 184, doi:110.1016/j.agee.2008.1006.1014, 2008.
- 810 Groupe Régional d'Action contre la Pollution Phytosanitaires des Eaux Lorraine GRAPPE Lorraine: Les produits phytosanitaires utilisés par l'agriculture lorraine en 2004/2005, in, 2005.

Hahn, C., Prasuhn, V., Stamm, C., Lazzarotto, P., Evangelou, M. W. H., and Schulin, R.: Prediction of dissolved reactive phosphorus losses from small agricultural catchments: calibration and validation of a parsimonious model, *Hydrological and Earth System Sciences*, 17, 3679-3693; doi:3610.5194/hess-3617-3679-2013, 2013.

815 Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950 - 2006, *Journal of Geophysical Research: Atmospheres*, 113, D20119, 2008.

Hirsch, R. M.: An evaluation of some record reconstruction techniques *Water Resources Research*, 15, 1781 - 1790, 1979.

Holvoet, K., van Griensven, A., Gevaert, V., Seuntjens, P., and Vanrolleghem, P. A.: Modifications to the SWAT code for modelling direct pesticide losses, *Environ. Model. Softw.*, 23, 72-81, 2008.

820 Honti, M., Scheidegger, A., and Stamm, C.: Importance of hydrological uncertainty assessment methods in climate change impact studies, *Hydrol. Earth Syst. Sci.*, 3301-3317, doi:3310.5194/hess-3318-3301-2014, 2014.

Honti, M., Rieckermann, J., Schuwirth, N., and Stamm, C.: Can integrative catchment management mitigate future water quality issues caused by climate change and socio-economic development?, *Hydrological and Earth System Sciences*, 21, 1593-1609, doi:10.5194/hess-21-1593-2017, 2017.

825 Honti, M., Bischoff, F., Moser, A., Stamm, C., and Fenner, K.: Micropollutant degradation in rivers: Suitability of field modelling and regulatory data, *Water Resources Research*, in preparation.

Jachner, S., Gerald van den Boogaart, K., and Petzoldt, T.: Statistical Methods for the Qualitative Assessment of Dynamic Models with Time Delay (R Package qualV), *Journal of Statistical Software*, 022, 2007.

830 Jackson-Blake, L. A., Dunn, S. M., Helliwell, R. C., Skeffington, R. A., Stutter, M. I., and Wade, A. J.: How well can we model stream phosphorus concentrations in agricultural catchments?, *Environ. Model. Softw.*, 64, 31-46, <http://doi.org/10.1016/j.envsoft.2014.11.002>, 2015.

Jackson-Blake, L. A., Sample, J. E., Wade, A. J., Helliwell, R. C., and Skeffington, R. A.: Are our dynamic water quality models too complex? A comparison of a new parsimonious phosphorus model, SimplyP, and INCA-P, *Water Resour. Res.*, 53, 5382-5399, 10.1002/2016WR020132, 2017.

835 Johnson, A. C., Ternes, T. A., Williams, R. J., and Sumpter, J. P.: Assessing the concentrations of polar organic microcontaminants from point sources in the aquatic environment: measure or model?, *Environ. Sci. Technol.*, 42, 5390 - 5399, 2008.

Jungnickel, C., Stock, F., Brandsch, T., and Ranke, J.: Risk assessment of biocides in roof paints, *Environmental Science and Pollution Research* 15, 258 - 265, 2008.

840 Kehrein, N., Berlekamp, J., and Klasmeyer, J.: Modeling the fate of down-the-drain chemicals in whole watersheds: New version of the GREAT-ER software, *Environ. Model. Softw.*, 64, 1-8, <http://doi.org/10.1016/j.envsoft.2014.10.018>, 2015.

Larsbo, M., Roulier, S., Stenemo, F., Kasteel, R., and Jarvis, N.: An Improved Dual-Permeability Model of Water Flow and Solute Transport in the Vadose Zone, *Vadose Zone Journal*, 4, 398-406, 10.2136/vzj2004.0137, 2005.

845 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, 2013.

Leip, A., Marchi, G., Koeble, R., Kempen, M., Britz, W., and Li, C.: Linking an economic model for European agriculture with a mechanistic model to estimate nitrogen losses from cropland soil in Europe, *Biogeosciences Discussions*, 4, 2215-2278, 2007.

850 Leu, C., Singer, H. P., Stamm, C., Müller, S. R., and Schwarzenbach, R. P.: Simultaneous assessment of sources, processes, and factors influencing herbicide losses to surface waters in a small agricultural catchment, *Environ. Sci. Technol.*, 38, 3827 - 3834, 2004a.

Leu, C., Singer, H. P., Stamm, C., Müller, S. R., and Schwarzenbach, R. P.: Variability of herbicide losses from 13 fields to surface water within a small catchment after a controlled herbicide application, *Environ. Sci. Technol.*, 38, 3835 - 3841, 2004b.

855 Leu, C., Schneider, M. K., and Stamm, C.: Estimating catchment vulnerability to diffuse herbicide losses from hydrograph statistics *J. Environ. Qual.*, 39, 1441 - 1450; doi:1410.2134/jeq2009.0323, 2010.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Transactions of the American Society of Agricultural and Biological Engineers*, 50, 885 - 900, 2007.

860 Moschet, C., Wittmer, W., Simovic, J., Junghans, M., Piazzoli, A., Singer, H., Stamm, C., Leu, C., and Hollender, J.: How a complete pesticide screening changes the assessment of surface water quality, *Environ. Sci. Technol.*, 48, 5423-5432, dx.doi.org/5410.1021/es500371t, 2014.

Müller, K., Bach, M., Hartmann, H., Spiteller, M., and Frede, H.-G.: Point- and nonpoint-source pesticide contamination in the Zwester Ohm catchment, Germany, *J. Environ. Qual.*, 31, 309 - 318, 2002.

865 Munz, N., Melo, L., Reyes, M., Schönenberger, U., Singer, H., Spycher, B., de Zwart, D., Junghans, M., Hollender, J., and Stamm, C.: Pesticides drive risk of micropollutants in wastewater-impacted streams during low flow conditions, *Wat. Res.*, 110, 366-377, 2017.

Office national de l'eau et des milieux aquatiques ONEMA: La banque nationale des ventes réalisées par les distributeurs de produits phytosanitaires (BNV-D), 2014.

870 Ort, C., Hollender, J., Schaerer, M., and Siegrist, H.: Model-based evaluation of reduction strategies for micropollutants from wastewater treatment plants in complex river networks, *Environ. Sci. Technol.*, 43, 3214-3220, 2009.

Parker, R., Arnold, J. G., Barrett, M., Burns, L., Carrubba, L., Neitsch, S. L., Snyder, N. J., and Srinivasan, R.: Evaluation of three watershed-scale pesticide environmental transport and fate models, *J. Am. Water Resour. Assoc.*, 43, 1424 - 1443 2007.

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Field Code Changed

Field Code Changed

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: French (Switzerland)

Formatted: English (U.K.)

- 875 Pullan, S. P., Whelan, M. J., Rettino, J., Filby, K., Eyre, S., and Holman, I. P.: Development and application of a catchment scale pesticide fate and transport model for use in drinking water risk assessment, *Sci. Total Environ.*, 563–564, 434–447, <https://doi.org/10.1016/j.scitotenv.2016.04.135>, 2016.
- R Core Team: R: A Language and environment for statistical computing, in, R Foundation for Statistical Computing, Vienna, Austria, 2017.
- 880 Reichert, P.: AQUASIM - a tool for simulation and data analysis of aquatic systems, *Water Science and Technology*, 30, 21–30, 1994.
- Renaud, F. G., Bellamy, P. H., and Brown, C. D.: Simulating pesticides in ditches to assess ecological risk (SPIDER): I. Model description, *Sci. Total Environ.*, 394, 112–123, 2008.
- Reusser, D. E., Blume, T., Schaeffli, B., and Zehe, E.: Analysing the temporal dynamics of model performance for hydrological models, *Hydrol. Earth Syst. Sci.*, 13, 999–1018, 10.5194/hess-13-999-2009, 2009.
- 885 Röpke, B., Bach, M., and Frede, H. G.: DRIPS - a decision support system estimating the quantity of diffuse pesticide pollution in German river basins, *Water Science and Technology*, 49, 149–156, 2004.
- Schwarzenbach, R. P., Escher, B. I., Fenner, K., Hofstetter, T. B., Johnson, C. A., von Gunten, U., and Wehrli, B.: The challenge of micropollutants in aquatic systems, *Science*, 313, 1072–1077, 10.1126/science.1127291, 2006.
- 890 Sideris, I., Gabella, M., Sassi, M., and Germann, U.: The CombiPrecip experience: development and operation of a real-time radar-rain-gauge combination scheme in Switzerland, 2014 International Weather Radar and Hydrology Symposium, 2014, 10 p.,
- Smith, E. P., and Rose, K. A.: Model goodness-of-fit analysis using regression and related techniques, *Ecological Modelling*, 77, 49–64, 1995.
- 895 Spycher, S., and Daniel, O.: Agrarumweltindikator Einsatz von Pflanzenschutzmitteln--Auswertungen von Daten der Zentralen Auswertung Agrarumweltindikatoren (ZA-AUI) der Jahre 2009 - 2010, Forschungsanstalt Agroscope Changins-Wädenswil ACW, 79 p., 2013.
- Stamm, C., Scheidegger, R., van der Voet, J., Singer, H., and Bader, H. P.: Organische Spurenstoffe im Rahmen von NADUF. Machbarkeitsstudie - Schlussbericht, Eawag, Dübendorf, 32 p., 2012.
- 900 Steffens, K., Jarvis, N., Lewan, E., Lindström, B., Kreuger, J., Kjellström, E., and Moeys, J.: Direct and indirect effects of climate change on herbicide leaching - A regional scale assessment in Sweden, *Sci. Total Environ.*, 514, 239–249, 10.1016/j.scitotenv.2014.12.049, 2015.
- Strahler, A. N.: Quantitative analysis of watershed geomorphology, *Transactions American Geophysical Union*, 38, 913–920, 1957.
- 905 Swiss Federal Statistical Office FSO: Census of agricultural enterprises (Landwirtschaftliche Betriebszählung) 2010, Neuchâtel, Switzerland, 2011.
- Swiss Federal Statistical Office FSO: Land use statistics (Arealstatistik) 2004/09, Neuchâtel, Switzerland NOLU04/20120821, 2012.
- Villamizar, M. L., and Brown, C. D.: A modelling framework to simulate river flow and pesticide loss via preferential flow at the catchment scale, *Catena*, 149, 120–130, 2017.
- 910 Vogt, J., Soille, P., De Jager, A., Rimaviciute, E., Mehl, W., Foisneau, S., Bodis, K., Dusart, J., Paracchini, M. L., Haastrup, P., and others: A pan-European river and catchment database, *European Commission, EUR*, 22920, 120, 2007.
- Wittmer, I. K., Bader, H.-P., Scheidegger, R., Singer, H., Lück, A., Hanke, I., Carlsson, C., and Stamm, C.: Significance of urban and agricultural land use for biocide and pesticide dynamics in surface waters, *Wat. Res.*, 44, 2850–2862, 2010.
- 915 Wittmer, I. W., Scheidegger, R., Stamm, C., Gujer, W., and Bader, H.-P.: Modelling biocide leaching from facades, *Wat. Res.*, 45, 3453–3460, 2011.

Field Code Changed

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

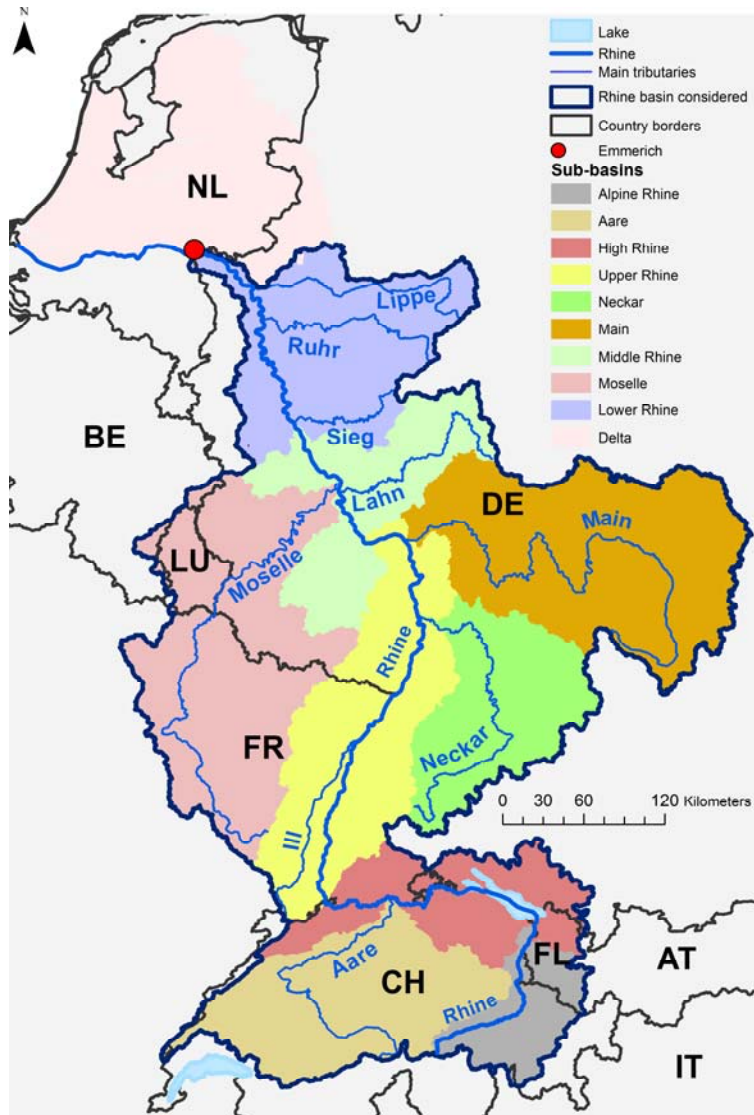
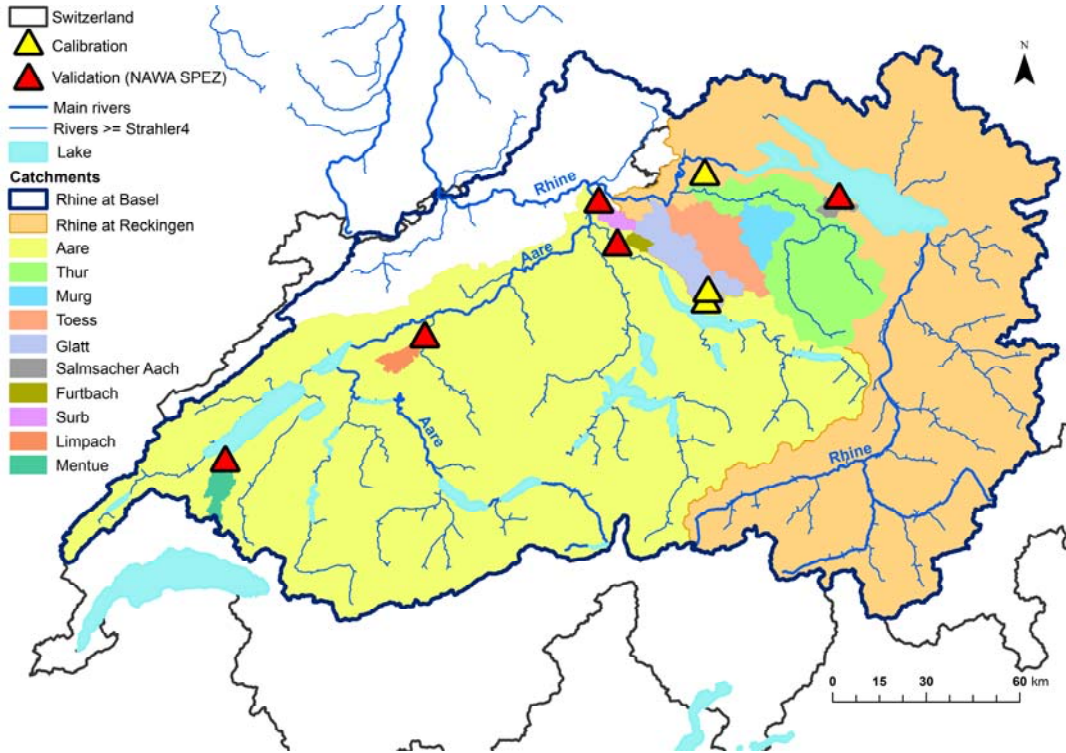


Figure 14: Map of the Rhine basin. The study area covers the part upstream of Emmerich indicated by the red circle. The different colours represent the sub-basins according to the International Commission for the Protection of the Rhine (ICPR) with the an additional distinction of the Aare basin in Switzerland. Base data: Vogt et al. (2007); Swisstopo (2007). [AT](#): Austria, [BE](#): Belgium, [CH](#): Switzerland, [DE](#): Germany, [FL](#): Liechtenstein, [FR](#): France, [IT](#): Italy, [LU](#): Luxembourg, [NL](#): The Netherlands.

Formatted: Font: 10 pt



930

Figure 22: Calibration and validation catchments in Switzerland. Base data: Vogt et al. (2007); Swisstopo (2007).

Formatted: Font: 10 pt, Not Bold

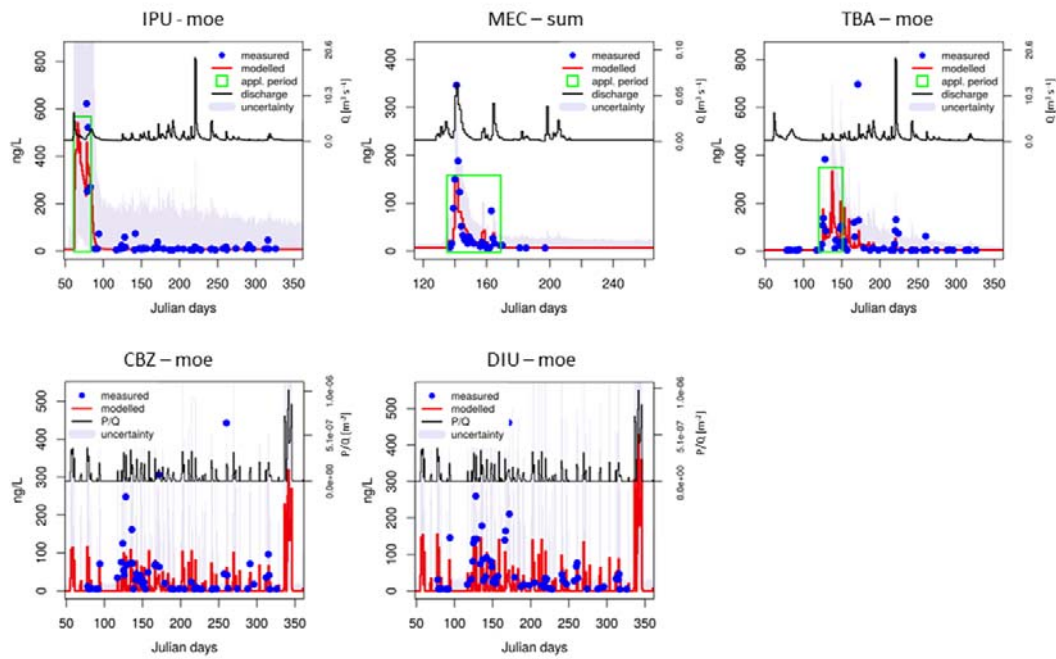


Figure 3: Examples of the comparison between simulated and observed concentration time series during the calibration step for each compound. IPU: isoproturon—(Mönchaldorf), MEC: S-metolachlor—(Summerau), TBA: terbuthylazine (Mönchaldorf), CBZ: carbendazim—(Mönchaldorf), DIU: diuron_Moe; (Mönchaldorf_sum; Summerau). The full set of calibrations is shown in the Supplementary Material (Fig. S7, S8S12).

935

940

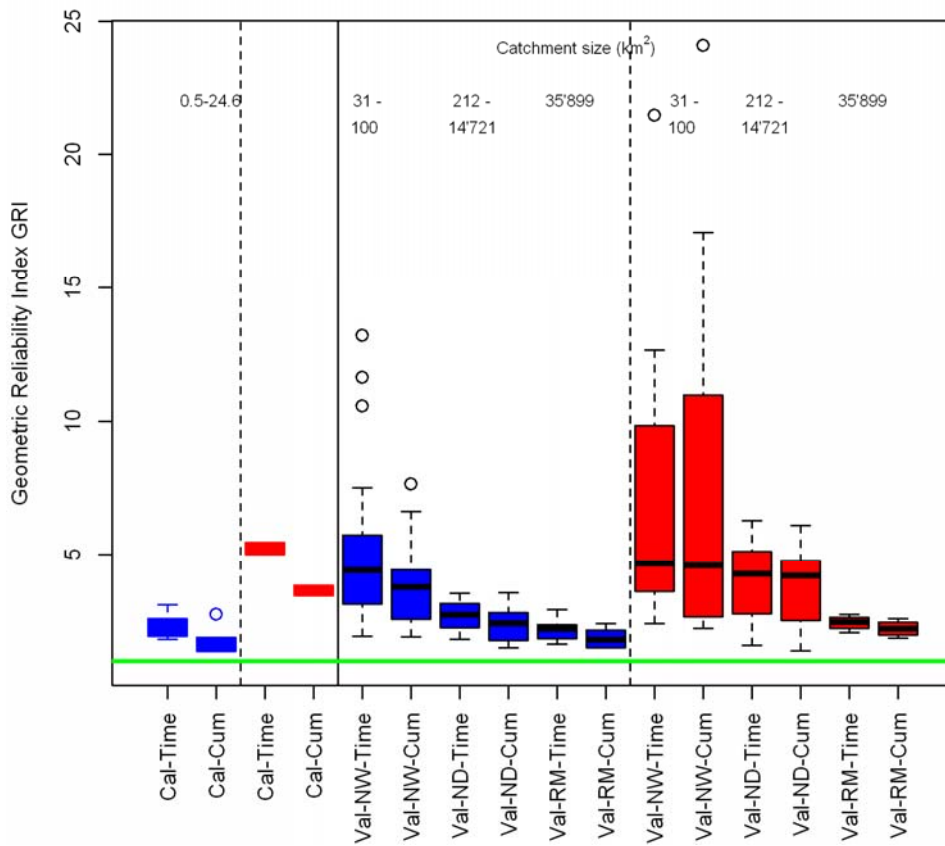


Figure 54: Overview about of the overall predictive power to simulate the concentrations levels during the calibration and validation phase as quantified by the geometric reliability index (GRI). A value of 1 (red-green horizontal line) indicates a perfect match; the larger the value the stronger the deviation. Cal: calibration; Val: validation; NW: NAWA SPEZ: National Surface Water Quality Monitoring Programme – Special Campaigns, ND: National long-term surveillance of Swiss rivers (NADUF), RM: International Rhine Monitoring Station (Basel), Time: evaluation of concentration time series; Cum: evaluation of cumulative concentration distributions (sorted according to size); green/blue: agricultural herbicides; orange/red: dual use (urban and agricultural) biocides.

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt

Formatted: Font: 10 pt, Not Bold

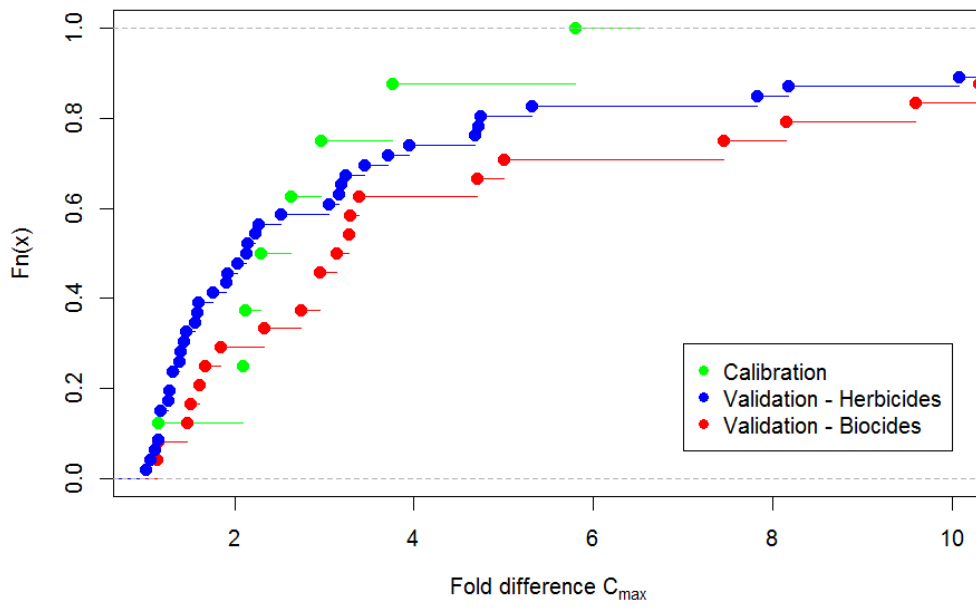


Figure 65: Cumulative distribution of the fold difference between observed and simulated concentrations C_{max} of all compounds during the calibration and validation phase.

Formatted: Font: 10 pt, Not Bold

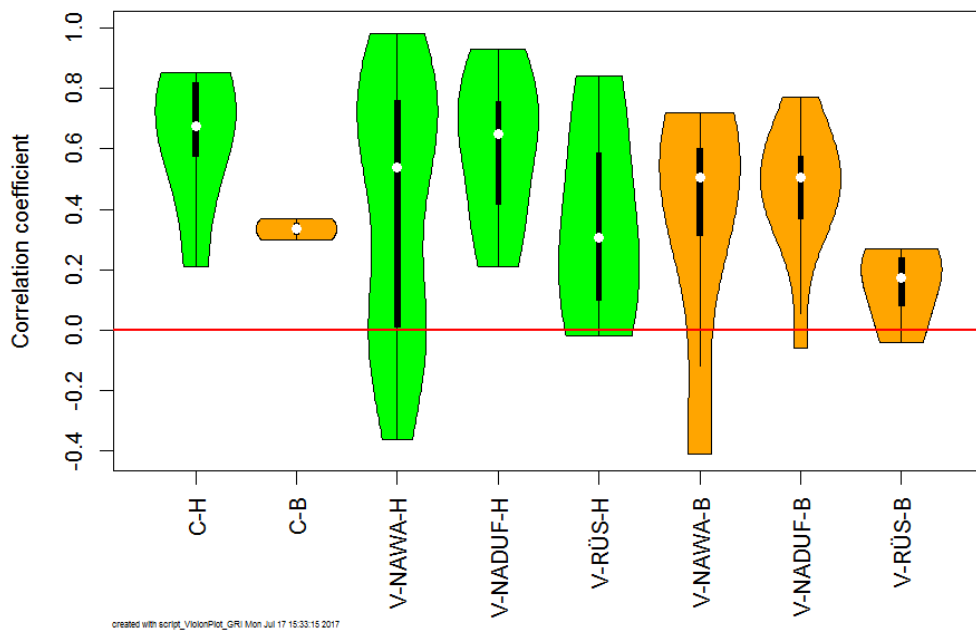
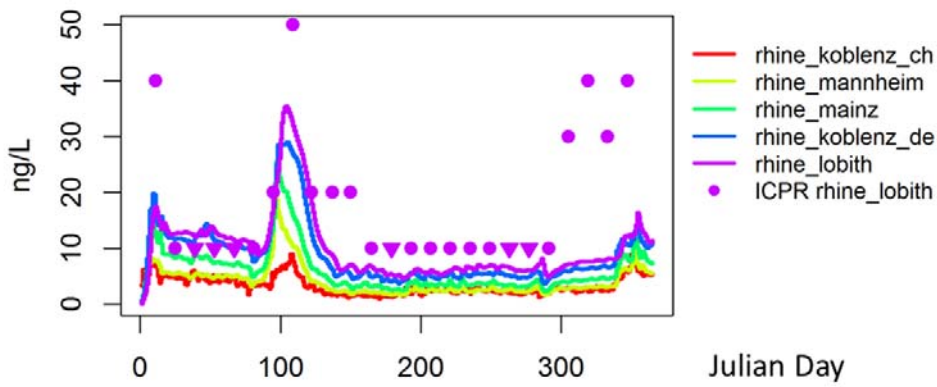


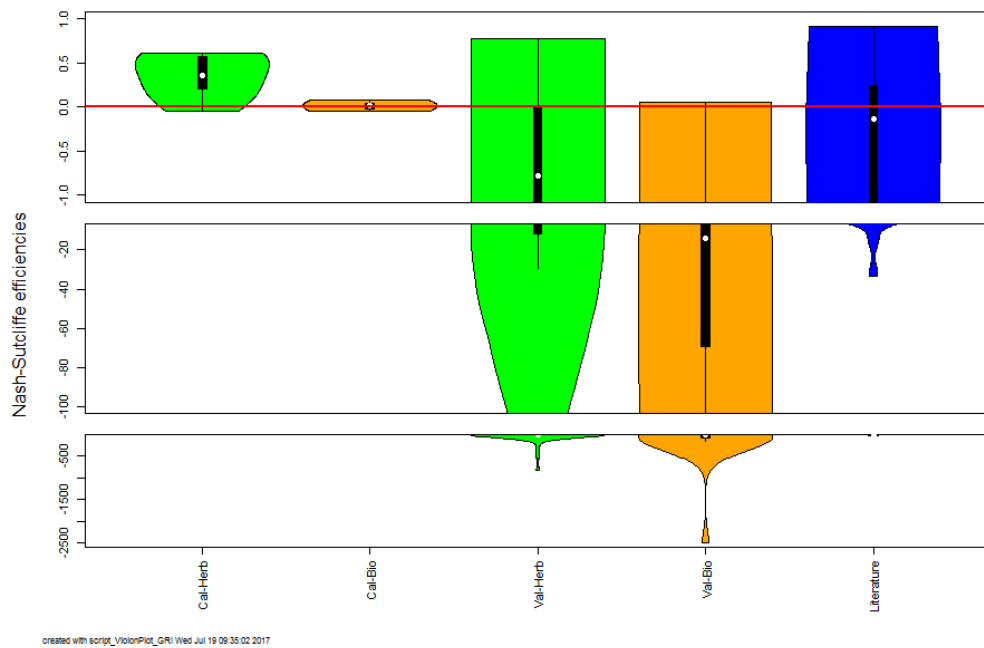
Figure 7: Violin-plots of the Pearson-correlation-coefficients between simulated and observed concentrations for (H) herbicides (green) and (B) biocides (orange) during (C) calibration and for the different (V) validation data sets.

960



965 **Figure 86:** Comparison of predicted isoproturon concentrations along the River Rhine for 2011 compared to the observations at the measuring site at Lobith.

Formatted: Font: 10 pt
 Formatted: Space After: 0 pt, Line spacing: 1.5 lines
 Formatted: Font: 10 pt, Not Bold



970 **Figure 9: Comparison of Nash-Sutcliffe efficiencies in this study for herbicides (green) and biocides (orange) during calibration (C) and validation (V) with values from the literature for diffuse pollutants (herbicides, different P-forms; sources: see Table 5).**

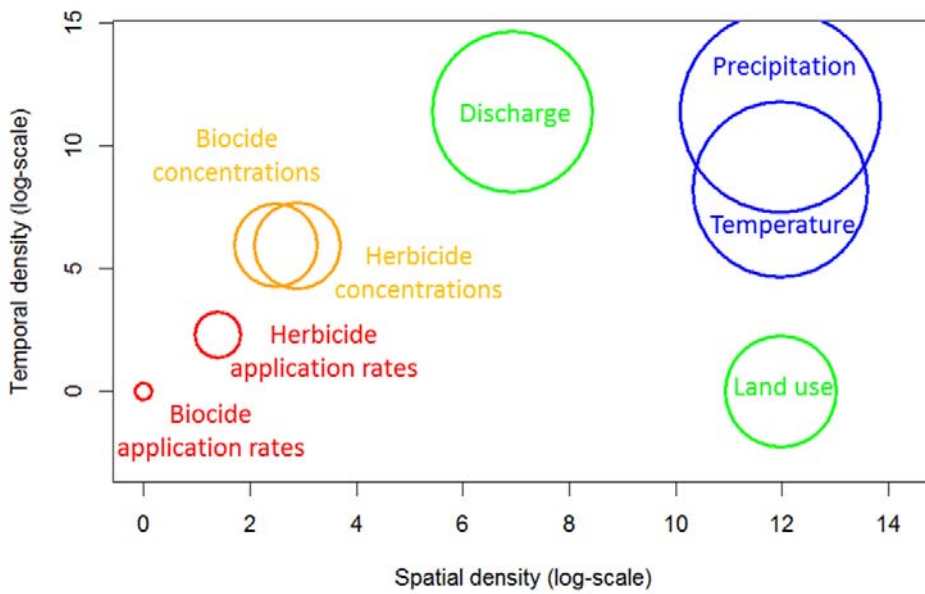
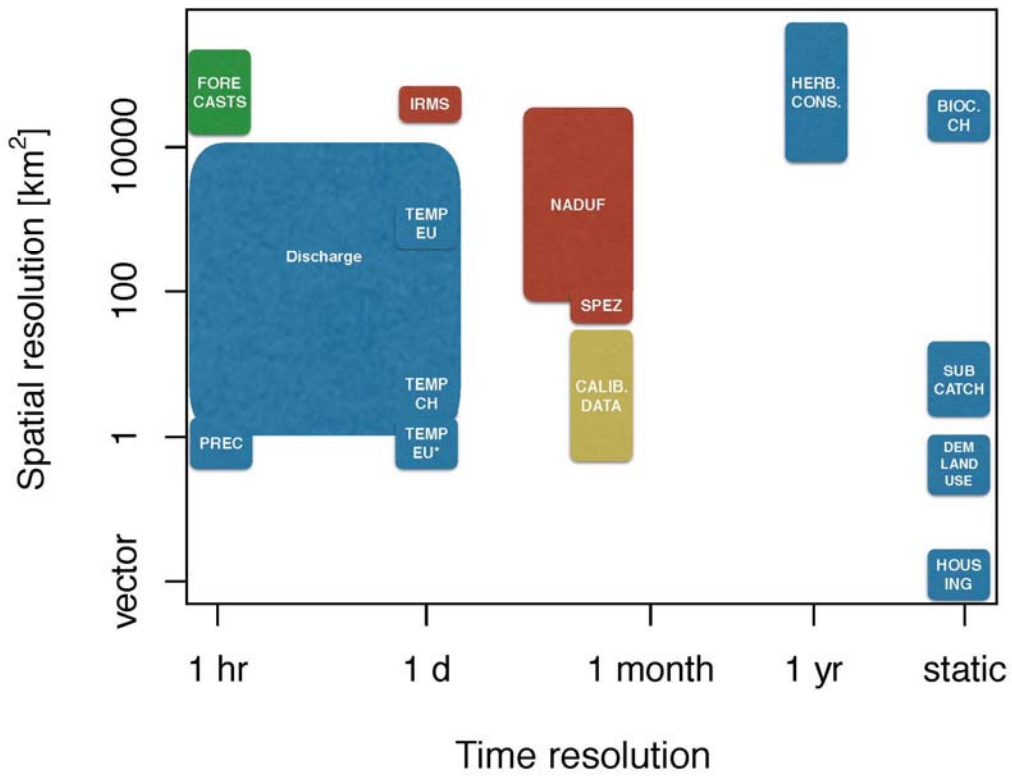
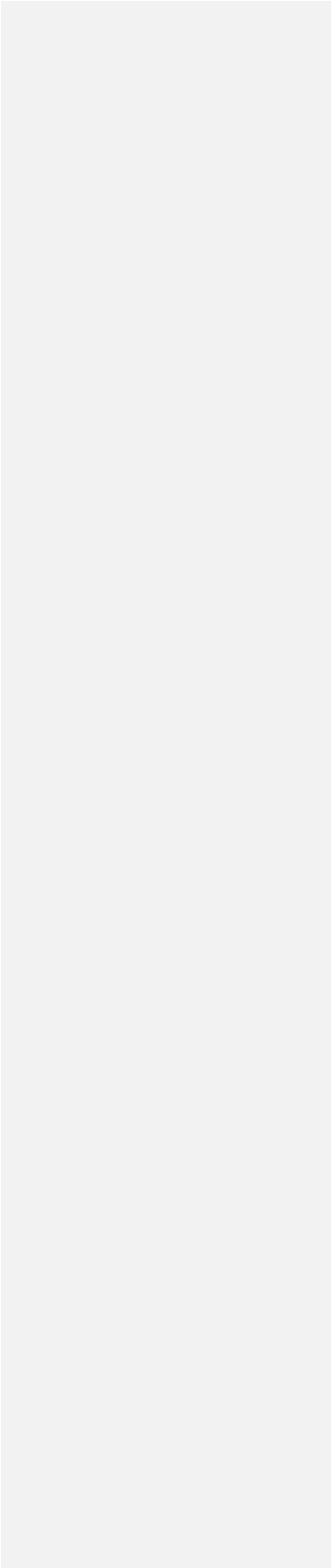


Figure 107: Spatial and temporal resolution of input variables (blue), calibration data (yellow), validation data (red), and forecasts (green). Abbreviations: prec - precipitation, temp - temperature, herb - herbicide, cons - consumption, bioc -

Formatted: Font: 10 pt, Not Bold

Formatted: Space After: 0 pt, Line spacing: 1.5 lines

biocides, catch - catchment Spatial and temporal density of different input variables. The different colours represent different categories of input quality.



Tables

985 Table 11: Global model parameters.

Abbreviation	Name	Description	Specificity	Model part
ρ	Initially available herbicide fraction	Fraction of the applied herbicide mass initially available for transport	compound	deterministic
k_{w-s}	Sorption rate	Fraction of the dissolved herbicide mass getting sorbed to the soil matrix per unit of time	compound	Deterministic model
k_{s-w}	Desorption rate	Fraction of the sorbed herbicide mass getting desorbed per unit of time	compound	Deterministic model
k_{deg}	Degradation rate	Rate constant of the first order degradation	compound	Deterministic model
C_{back}	Background concentration	Constant background concentration, proportional to the areal fraction of the relevant crop in the subcatchments	compound	Deterministic model
ϵ	Herbicide loss rate	Loss rate per unit discharge and available herbicide mass	compound	Deterministic model
T_{obj}	Temperature objective	Cumulative temperature sum required to start herbicide application on a crop	crop	Deterministic model
β	Biocide loss rate	Loss rate per unit precipitation and available biocide mass	compound	Deterministic model
μ	Scaling factor	Factor for scaling the model error term proportional to the subcatchment-specific herbicide input	compound	Error model
σ_{error}	Standard deviation of the error model	Relative standard deviation of the total model error	compound	Error model

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Table 22: Characterization of calibration and validation catchments. NADUF: National long-term surveillance of Swiss rivers. NAWA SPEZ: NAWA SPEZ: National Surface Water Quality Monitoring Programme – Special Campaigns. IRMS: International Rhine Monitoring Station (Basel).

Catchment/ River	Abbr.	Reference	Year	Area [km ²]	agricultural land use [km ²]	Housing footprint [km ²]	population
Calibration							
Ossingen	oss	Doppler et al. 2012	2009	1.2	1.1	-	-
Summerau	sum	Gomides Freitas et al. 2008	2003	0.5	0.04	-	-
Mönchaltorf	moe	Wittmer et al. 2010	2007	24.6	4.7	0.5	12'000
Validation							
Furtbach	fch	NAWA SPEZ	2012	31	14	1.6	31'570
Limpach	lch	NAWA SPEZ	2012	74	43	1	7'560
Mentue	mnt	NAWA SPEZ	2012	100	42	1	9'300
Salmsacher							
Aach	smr	NAWA SPEZ	2012	54	33	1.7	17'326
Surb	srb	NAWA SPEZ	2012	68	36	1.4	22'780
Thur	thr	NADUF	2009	1'735	873	33	403'028
Toess	tss	NADUF	2009	432	175	11	197'032
Glatt	glt	NADUF	2009	413	183	20	405'702
Murg	mrg	NADUF	2009	212	118	5.3	68'145
Rhine-							
Reckingen	rhn	NADUF	2009	14'721	5'261	175	2'946'907
Rhine-Basel	irms	IRMS	2010/11	35'899	12'009	503	7'786'398

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Formatted: Font: Not Bold

Formatted Table

Table 33: Metrics used for quantifying model performance. C_{max}^{obs} : observed maximum concentration, C_{max}^{sim} : simulated maximum concentration, m_i : model prediction, \bar{m} : mean model prediction, n : number of observations, o_i : observation, \bar{o} : mean of the observations, σ_{obs} : standard deviation of the observations.

Metric	Abbreviation	Description
Nash-Sutcliffe Efficiency	NSE	$NSE = 1 - \frac{\sum_{i=1}^n (o_i - m_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2}$
Pearson correlation coefficient	r	$r = \frac{\sum_{i=1}^n (o_i - \bar{o})(m_i - \bar{m})}{\sqrt{\sum_{i=1}^n (o_i - \bar{o})^2} \sqrt{\sum_{i=1}^n (m_i - \bar{m})^2}}$
Percent bias	PBIAS	$PBIAS = 100 \times \frac{\sum_{i=1}^n (m_i - o_i)}{\sum_{i=1}^n o_i}$
Relative root mean square error	RRMSE	$RRMSE = \frac{\sum_{i=1}^n m_i - o_i }{n \sigma_{obs}}$
Geometric Reliability Index (cumulative distribution)	GRI (GRI_sorted)	$GRI = \frac{1 + \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{m_i - o_i}{m_i + o_i} \right)^2}}{1 - \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{m_i - o_i}{m_i + o_i} \right)^2}}$
Relative difference between maximum concentration	$\Delta_{C_{max}}$	$\Delta_{C_{max}} = \frac{C_{max}^{sim} - C_{max}^{obs}}{C_{max}^{obs}}$
Fold difference between maximum concentration	F.diff	$F.diff = \begin{cases} \frac{C_{max}^{sim} - C_{max}^{obs}}{C_{max}^{obs}} & C_{max}^{sim} > C_{max}^{obs} \\ \frac{C_{max}^{obs}}{C_{max}^{sim} - C_{max}^{obs}} & C_{max}^{sim} < C_{max}^{obs} \end{cases}$

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

1000

Table 44: Over- or underestimation of maximum concentrations (site-compound combinations) in percentage of the observations. For the herbicides only the peaks during spring application were considered. IPU: isoproturon, MEC: S-metolachlor, TBA: terbuthylazine, CBZ: carbendazim, DIU: diuron. NADUF: National long-term surveillance of Swiss rivers, NAWA SPEZ: NAWA SPEZ: National Surface Water Quality Monitoring Programme – Special Campaigns, IRMS: International Rhine Monitoring Station (Basel).

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

		IPU	MEC	TBA	CBZ	DIU
Calibration	Mönchaltorf	-13	-51	-53	-62	-66
	Ossingen	-71	-	-83	-	-
	Summerau	-	-58	-	-	-
Validation <u>NAWA</u>						
SPEZ	Furtbach	6	-10	431	61	715
	Salmsacher					
	Aach	114	17	1898	229	1201
	Surb	123	-53	56	-32	859
	Limpach	103	-14	17	-57	2772
	Mentue	2405	45	43	84	370
Validation NADUF	Thur	-9	-47	-	-57	91
	Rhine					
	Reckingen	22	20	-	-65	70
	Murg	-42	-61	-	-97	221
	Toess	-35	-37	-	458	265
	Glatt	92	-45	-	4	789
Validation IRMS	Rhine Basel	-67	-60	368	239	931

1005

Table 55: Examples of reported Nash-Sutcliffe efficiency values and Pearson correlation coefficients between observations and simulations reported for a selection of water quality modelling studies. C: calibration, V: validation.

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Reference	Compound	C/V	NSE	r
(Bannwarth et al., 2014)	Atrazin	C	0.92	-
		V	0.61	-
	Chlorothalonil	C	0.67	-
		V	0.28	-
	Endosulfan	C	0.86	-
		V	0.31	-
(Parker et al., 2007)	Atrazine	C	-0.18/-1.03/-3.50 [†]	0.12/0.30/0.64 [†]
	Metolachlor	C	-0.84/-3.53/-33.4 [†]	0.14/0.46/0.57 [†]
	Trifluralin	C	-30.2/-16.9/-3.2 [†]	-0.16/0.35/0.14 [†]
(Boulangé et al., 2014)	Mefenacet	S	0.65/-9.72/-14.7 [‡]	0.78/0.87/0.92 [‡]
(Holvoet et al., 2008)	Atrazine	C	0.66	-
Holvoet 2007	Chloridazon	C	-0.67 [#]	0.44 [#]
(Jackson-Blake et al., 2015)	Suspended sediment	C	0.16/0.39/0.21/0.02 [*]	0.63/0.83/0.64/0.21 [*]
	TDP	C	0.24/0.04/-0.20/-0.60 [*]	0.83/0.68/-0.05/0.27 [*]
	Different P forms	C	0.06/-0.14/-0.60/-0.42/- 1.15/-4.18/0.19/-0.08/- 0.74/0.08 ^S	

[†]: values for three different models