We would like to thank Adèle Bressy for the time she has taken to read our manuscript and her helpful comments to improve it. In the following section we are going to repeat the points brought up (in grey italic letters) and subsequently respond to them:

Specific comments

Title: Perhaps specify in the title that it deals with stormwater –

We will change the title to "Sources and pathways of biocides and their transformation products in urban stormwater infrastructures of a 2 ha urban district".

Abstract: The last sentence of the abstract present obvious conclusions and not informative. It seems logical that by sampling in a targeted way, a better identification of the sources is obtained. Perhaps you can refine this conclusion and add some more concrete and precise results.

Thank you for your suggestion. We will change the last sentence: The applied two-step approach determined sources and pathways of biocide and their TPs. This study contributes to expanding knowledge on their entry and distribution and thus eventually towards reducing emissions.

Line 57: Please explain the choice of the TPs, why just these 3 compounds?

These three compounds are commonly used as film protection products. They represent one herbicide, one algicide and one fungicide. Often, a combination of these and more compounds is used against algae and fungi growth (Sauer, 2017). All three compounds and the selected TPs have been part of previous studies on biocide runoff from facades, e.g. Burkhardt et al., 2011, Bollmann et al., 2016, Bollmann et al., 2017, Hensen et al., 2018, Paijens et al., 2021. For the quantification of TPs, an analytical standard needs to be available. Standards were available for the selected TPs. We will clarify our choice in an updated manuscript accordingly.

Paragraph 2.1: the tow-step approach is well presented and convincing, but the long period between the first campaign (step 1 in 2015-2017) et the last one (step 2 in 2019-2020) raises the question of the comparability of the campaigns between them. Why did you not sample the swale system during the second campaign in 2019-2020 to verify the stability of the concentrations in the swale? Justify this point.

Thanks for this comment. On a first glance, it really seemed obvious to continue swale sampling also in 2019-2020. However, as shown in Figure 4, biocide concentrations in the swale are highly variable and depended inter alia on event magnitude. Hence, we did not expect new findings from a renewed sampling campaign here. Instead, our objective was to concentrate on biocide sources and thereby limit the number of samples in an efficient campaign. We will stress this point in an updated manuscript.

Line 117: you said that the last paint was in 2007 and after it is indicated that a façade was painted in 2018 (line 223). It is unclear.

All buildings were painted last in 2007. However, there is one part of a façade that was re-painted due to restauration works in 2018. We will clarify this.

Line 121: Is there always water in the swale or is it dry during dry weather?

The swale is episodic, i.e. dry during dry weather. We will add this information.

Line 153: how are sampled the roof, façade and pipe samples? Are they representative of the entire rain events? What about the first flush? You should add details about the sampling and its representativeness.

All pipe samples (downpipes, street, drainage) were point samples during rain events and did not include first flush effects that might have shown higher concentrations. They are also not

representative of the entire rain event as no flow proportional samples were taken (see 3.4). We are aware of the concentration distribution of biocides during rain events and also of the first flush e.g. Bollmann et al., 2014. We will discuss this point in more detail in the manuscript.
Roof samples and facades samples were taken during artificial elution experiments as described in 2.3.2. We will rewrite this paragraph to describe the sampling in more detail.

Line 153: how the water is sampled? You said during the sample, why not at the end of the sample?
Water samples at rain downpipes were taken during the event. We chose to sample every event only once during the event because we had limited analytical capacity. See also comment above.

Line 163-165: did you analyse the representativeness of the sampled events in relation to the classical pluviometry?
Thank you for this idea. So far, we did not analyze the representativeness of the sampled events. In principle, sampling was only possible when there was water in the swale, which produced a bias towards large events. As shown in Fig. 4, we took samples during 3 of the 5 largest events in our measurement period. All sampled events were larger than 4mm/day. We will include a comparison to longer-term rainfall data and recurrence intervals in the updated manuscript. However, we will limit the validity of this analysis, since, as already described, the weather station is 5km away from the study area and there might be differences in local precipitation.

Figure 4: I am not sure that this figure is really informative since we are not able to read the rainfall for each sampled event. Perhaps put it in supplementary materials
This figure aims to show an overview of the chronology of the sampling. We will add the rainfall amount to the sampled events; see also comments to the referee #1 and #2.

Line 176: why did you not test solar panel elution? Do you think that they could emit biocides?
We are not aware of studies that found biocides used in solar panels, especially Diuron, Terbutryn and OIT measured in this study. We found very low Terbutryn concentrations and concentration of measured TPs in rain downpipes of houses with solar panels (Fig. 7). For this reason, we did not look for sources. We will clarify this point in the discussion.

Line 177-178: n=1 seems insufficient to conclude.
The limited significance of these samples will be discussed in the updated manuscript.

Chemical analysis: I would recommend to present the analytical validations (as extraction recoveries) and the analytical uncertainties to validate the SPE extractions and the quantification.
Thank you for your suggestion. We will add the extraction recoveries here. Recovery was determined by spiking water samples with 1 mg $L^{-1}$of analytical standard and was found to be 97.7 % (Diuron), 88.5 % (Terbutryn) and 93.5 % (OIT), 85.0 % (Diuron-desmethyl), 66.2 % (Terbumeton), 50 % (Terbuthylazin-2-hydroxy) and 92 % (Terbutryn-desethyl) (Hensen et al., 2018).

Paragraph 2.5: I am not really convinced by the methodology presented by this paragraph because the concentration used does not take into account the temporal evolution of emitted concentrations over time due to ageing or depletion of the stock in the material, or does not present an argument from the literature to overcome this. What verification have you implemented to justify the word "efficiently" in line 222. Justify the use of an average biocide concentrations to calculate BE. Moreover, why the used samples were not described in the 2.3 section? What is the number of the samples and the representativity? What is the sampling frequency? What is the variability of the measured concentrations? Does the concentration vary in time? Decrease? To prevent the reader's doubts, a part of the explanation from line 349 to 355 could be used in

the methodology presentation and the fact that the estimated BE will be compared to the literature.

Thank you for this comment and your suggestions. We are aware that our approach is limited and only a rough estimate. We chose the word "efficiently" to stress that by only very few samples and little information on the building we obtained realistic estimations on biocide emissions over a two year time period. Sampling is described in 2.3.1 as part of sampling the rain downpipes. We describe concentrations in section 3.2.2. They vary for different events and rather decrease over time which compares to expectations. But the exact temporal evolution of concentrations cannot be determined based on only four and five point samples of events. We will modify this paragraph to clarify the limitation of our approach already in the method section.

Paragraph 3.1: I am wondering if 4 sampled events are sufficient to assess the variability of the concentrations in the swale, especially since only one PNEC exceedance is observed to justify the continuation of the study. Why don't you continue to sample the swale in the second part of the study?

See our answer to paragraph 2.1 above.

Line 298: You said that you sampled an additional pipe (R4-2) because R4-1 exceeded R1 and R2 by an order of magnitude but you have no result for R1 and R2 before the first sampling of R4? I don't understand.

Your observation is correct, thank you for finding this contradiction. We first sampled one pipe at one building (R4-1) and found concentrations that exceeded expected concentrations, because we did not expect roof areas as a biocide source. During the next event, we decided to sample another pipe at the same building (R4-2) to make sure there was no contamination in the first pipe. Additionally, we decided to sample one pipe at another house (R2). To confirm the low concentrations we sampled at a third building (R1). For comparison purposes we then sampled multiple events at all pipes. We will make this clear in an updated manuscript.

Figure 8: Precise if it is mean or median values in the legend

Shown are mean values. We will add this to the figure description.

Line 320 and following: It is not clear if the difference of concentrations is due to the new paint or to the exposition.

We will change this sentence to make it clear that the different concentrations are due to the new paint.

Line 363/364: you explain that OIT was not detected due to its degradation in soil but for S9 (surface water pipe), water is not percolated through the soil? How do you explain to not found OIT in S9 samples?

Concentrations of OIT at the facades were very low, i.e. 0.9-2.3ng/L. Hence, we did not expect to find OIT in the pipes. We will add this point to the manuscript.

Line 413: you have to qualify this sentence because your method gave a rough and short term estimate (even if interesting). the comparison to the literature is only informative and does not bring proof of the accuracy of the evaluation, the initial stock of biocides not being necessarily the same.

Thank you for pointing this out. We will clarify this statement. Samples at the rain down pipes just confirmed a continuous biocide leaching from the flat roof. Based on available measurements we did a rough estimation of the total long-term biocide leaching and a comparison of the obtained estimation with literature values to check, if we arrived at a realistic order of magnitude. We will clarify this both in the method and in the result section, see also our response to the comment on paragraph 2.5 above.

Technical corrections

Figure 1: Step 2: perhaps precise "Elution and leaching test experiments" - will be corrected.

Line 97: two -s at "sselected" - will be corrected.

Line 114: The capital letter at "Area" is unnecessary - will be corrected.

Line 118: perhaps add a -s at "diverse use"? - will be corrected.

Line 134: I find that "surface water pipe" do not describe well the type of water sampled. It looks like surface water that has been sampled. Perhaps "surface runoff pipe" would be more meaningful

Thank you for your suggestion, we will change the term accordingly.

Figure 5: you could cut the ordinate-axis to better present the lowest concentrations.

Thank you for your suggestion, we will change the figure accordingly.

Line 246: space is missing between below and 4 - will be corrected.

Line 245, 246 and 247: the sentence is not simple to understand

We will change this sentence to make it clearer.

*Line 254: substanceS -*will be corrected.

*Line 296: perhaps add a coma after "in all rain downpipes" -*will be corrected.

Figure 7 (d): perhaps precise "non sampled event". The use of one single scale is understandable but does not allow to read the concentrations for R1 and R2

We will add non sampled events in the explanation of the figure.

*Line 321: due "to"? -*will be corrected.

Line 579: "TEXTE"? -will be changed.

References

Bollmann, U. E., Fernández-Calviño, D., Brandt, K. K., Storgaard, M. S., Sanderson, H., and Bester, K.: Biocide Runoff from Building Facades: Degradation Kinetics in Soil, Environmental science & technology, 51, 3694–3702, doi:10.1021/acs.est.6b05512, 2017.

Bollmann, U. E., Minelgaite, G., Schlüsener, M., Ternes, T., Vollertsen, J., and Bester, K.: Leaching of Terbutryn and Its Photodegradation Products from Artificial Walls under Natural Weather Conditions, Environmental science & technology, 50, 4289–4295, doi:10.1021/acs.est.5b05825, 2016.

Bollmann, U. E., Vollertsen, J., Carmeliet, J., and Bester, K.: Dynamics of biocide emissions from buildings in a suburban stormwater catchment - concentrations, mass loads and emission processes, Water research, 56, 66–76, doi:10.1016/j.watres.2014.02.033, 2014.

Burkhardt, M., Zuleeg, S., Vonbank, R., Schmid, P., Hean, S., Lamani, X., Bester, K., and Boller, M.: Leaching of additives from construction materials to urban storm water runoff, Water Science and Technology, 63, 1974–1982, doi:10.2166/wst.2011.128, 2011.

Hensen, B., Lange, J., Jackisch, N., Zieger, F., Olsson, O., and Kümmerer, K.: Entry of biocides and their transformation products into groundwater via urban stormwater infiltration systems, Water research, 144, 413–423, doi:10.1016/j.watres.2018.07.046, 2018.

Paijens, C., Bressy, A., Frère, B., Tedoldi, D., Mailler, R., Rocher, V., Neveu, P., and Moilleron, R.: Urban pathways of biocides towards surface waters during dry and wet weathers: Assessment at the Paris conurbation scale, Journal of hazardous materials, 402, 123765, doi:10.1016/j.jhazmat.2020.123765, 2021.

Sauer, F.: Microbicides in Coatings, 143 pp., 2017.