Author's response

We thank all the reviewers for taking the time to evaluate our manuscript and for the thoughtful comments and efforts to provide constructive feedback, which was invaluable in helping to improve the quality and clarity of our work. We have carefully considered each of the comments and have addressed them below. Any references to line numbers made in our response relate to those in the original manuscript. In addition to the revisions made to accommodate the reviewer suggestions, we wanted to point out that also most of the values in the tables in the results section, and the use of these numbers in the text, changed slightly. Based on some of the reviewer comments, we decided to update the urban extent grid to a more recent dataset (from 2009 to 2015). While this change slightly rearranged the contaminant pathways, and thus the related numbers in the tables, it did not cause any substantial changes in the overall results and conclusions of the manuscript. Rather, it lead to a more accurate representation of the underlying processes in the model.

Response to Reviewer 1

R1-C1

Regarding the lower calculated concentrations in river, the authors have indicated that river discharge may be higher than observed and veterinary and industrial are not considered. I agree on this point. But on the other hand, it should be mentioned that there are factors that further decrease the river concentration. Within this paper, there is no mention of load reduction before entering the wastewater treatment plant. If the direct discharge coefficient is considered as the inflow from the conduit to the river in urban untreated area, I believe that something similar may be happen in the sewer pipes. Taking this into account will lead to a decrease of river concentration. In addition to this, advanced wastewater treatment plants could lead to further load reductions. It would be desirable to mention these points in the discussion or as future research topics.

A: We agree with the reviewer that sewer leakage, environmental decay and other losses before the contaminant load enters wastewater treatment plants could substantially reduce the river concentration. I.e., if we added these processes to our model, our simulations would lead to even lower calculated concentrations, hence our bias would be further amplified. That said, we added a new paragraph after line 596 to discuss these additional uncertainties:

"The efficiency of a WWTP to remove a specific contaminant is also a complex process that depends on characteristics of the individual facilities and local conditions that are not represented in the global HydroWASTE database. Furthermore, processes not simulated by HydroFATE may have an impact on contaminant loads entering surface waters. For example, depending on how far a household is located from the facility, decay processes in sewers can act on the load on its way to the WWTP. In addition, sewer lines that are poorly maintained may result in wastewater leakages into the ground, further reducing the load of contaminant before it reaches the WWTP."

Also, we changed the statement at line 649:

"As it is not possible to isolate the contribution from domestic sources in the MECs, this uncertainty in both PECs and MECs could explain a portion of the high negative bias found in the evaluation."

We also agree that WWTPs with an advanced level of treatment would typically lead to higher load reductions, thus HydroFATE has the capability to differentiate efficiencies from different levels of treatment. Our decision to use the same efficiency for all treatment levels in our test case study was due to the lack of data on this parameter for the chosen substance sulfamethoxazole. This uncertainty is discussed in lines 635-640, but we added a new statement at line 640:

"This could lead to reduced river concentrations, since secondary and advanced treatment processes are expected to result in higher removal efficiencies."

Minor comments:

R1-C2

It would be good to indicate the 10 km on line 147, if there is any reasoning behind it.

A: According to the methodology used to estimate the outfall location of the wastewater discharge location in the global WWTP database HydroWASTE, 10 km is the maximum distance between the actual facility and the outfall location and it is a described uncertainty in HydroWASTE. The distance was selected based on a statistical determination process using a subset of WWTPs and remote sensing imagery for manual verification (see Ehalt Macedo et al., 2022). To clarify this in the manuscript we added in line 148:

"…, given the locational uncertainties in HydroWASTE of up to 10 km (Ehalt Macedo et al., 2022),…"

R1-C3

Isn't ds,r in line 321 a mistake for dl,r?

A: We thank the reviewer for spotting this typo, though the error is actually in line 322 rather than in the equation. We corrected line 322 to "…$d_{l,r}$ (dimensionless) is the lake decay factor …".

R1-C4

Check line 332 for a reference error.

A: The reference error in line 332 has been reviewed and corrected.

R1-C5

There is a spelling error (individual) in 533 in Fig. 5.

A: The spelling error "individual" in the caption of Figure 5 (line 533) has been fixed.

Response to Reviewer 2

Critical:

R2-C1

According to GMD code and data availability policy: "Where the authors cannot, for reasons beyond their control, publicly archive part or all of the code and data associated with a paper, they must clearly state the restrictions" (https://doi.org/10.5194/gmd-12-2215-2019). In the supplement python file, HydroFATE_v09.py, the imported library arcpy is not publicly available. Please state the restrictions in the code and data availability section. For example: "A license of the software provided by ... is required to run the provided scripts."

A: We acknowledge the important issue about the arcpy library not being publicly available. We updated the code and data availability section to clearly state this restriction at line 667:

"*A license of the software ArcGIS Pro provided by ESRI is required to run the provided scripts.*"

R2-C2

The method presented in Section S.1 of the supplementary material is crucial to reproducing the modeling result of HydroFATE v1.0. In essence, Section S.1 details the generation algorithm of the WWTP service area. Section S.1 should be moved to the main manuscript, perhaps in the Appendix.

A: We agree that the generation algorithm of the WWTP service area (Section S.1) is crucial to reproducing our modeling results (we originally had placed the descriptions into the SI only for space considerations). We now moved this section to the main manuscript as an appendix, to make it more accessible to readers.

R2-C3

Follow up on the previous comment on the generation of the WWTP service area. Line 242, which is in Section 3.1, reported a "successive trial-and-error approach in which intermediate results were mapped, visually inspected for plausibility, and statistically tested to verify whether they led to further improvements." I understand the difficulty of the service area generation, and it is acceptable to inspect the model visually. However, the realization of the WWTP service

area is a critical part of the HydroFATE model. Hence, it is within the scope of the GMD code and data availability policy. I urge the authors to provide the scripts you use to generate the service area since I cannot find them in HydroFATE_v09.py. (Please review Section 3.2 of the GMD editorial: https://doi.org/10.5194/gmd-12-2215-2019.) Furthermore, the sensitivity analysis in Section S.3 does not provide the sensitivity of different realizations of the WWTP service area. Therefore, it is necessary to make the service area generation process more transparent by providing the script.

A: We thank the reviewer for this comment. We understand the concern about the transparency of the WWTP service area generation process. To address this, in the revised version of the paper we will provide the script used to generate the WWTP service areas, as suggested in Section 3.2 of the GMD editorial. This will enhance the reproducibility and transparency of our work. As such, we added the following sentence at line 659:

"The code for the WWTP delineation and its output is also available under at the same URL."

Suggestions:

R2-C4

Figure 2 provides an example of the generated WWTP service area used in this study. It would be helpful to show examples of intermediate iteration results and explain why this iteration is rejected or accepted so the readers can have a gist of the visually inspected acceptance criteria.

A: We have thoroughly considered this interesting comment. While we understand the value of a more comprehensive representation of the iterative steps involved, we struggled in finding representative in-between stages that are easy to explain yet illustrative and representative for the broad range of issues that we encountered. We felt that either the figure and explanations would get overly complex, or the simplifications may cause further confusion for the reader. To stay within reasonable limitations in terms of space and detail in this publication, we therefore hope that including the script and the output in the revised paper (see previous comment and answer) will be sufficient to help readers who have an interest in applying our methodology in their own work, or to experiment with modifications to the algorithms to study the associated effects.

R2-C5

In equation (1), the removal efficiencies of wastewater treatment plants (WWTP), $e_{WWTP,\ j}$, contain three treatment levels, which are primary, secondary, and advanced. However, the parameters of the case study do not include such complexity. Only one removal efficiency is used in each scenario. During the review, I have always considered WWTP treatment levels spatially varying, and they can be sensitive to the modeling result. I find the presentation of Figure 2 a bit misleading. Therefore, I suggest adding another figure in Section 4, which is based on Figure 2, but the WWTP areas are only color-coded once to represent the model better.

A: We understand and appreciate this concern. Our goal is not to mislead the reader by presenting different WWTP types in Figure 2. But we would like to emphasize that our case study, presented in Section 4, serves only as a test application of the model. In that case, regrettably, certain parameters essential for the selected substance were unavailable for inclusion in the study and some aspects of the analysis presented in Section 4 may not fully capture the complete scope of the model's capabilities. Therefore, we believe that adding another, simplified figure would not do justice to presenting the comprehensive and accurate features of the model. The model has explicitly been developed to be able to distinguish different WWTP types, it is only in the test study where we could not implement this distinction. We therefore hope that it is acceptable to leave the presentation of Section 4 as is. We would also like to point to Figure 4 which presents the WWTPs used in India, all visualized without their specific type.

Minor:

R1-C6

The unit in Figure 4 should be presented as $m^3\ s^{-1}$.

A: As requested, we corrected the units in the caption of Figure 4 to be presented as $m^3\ s^{-1}$.

R1-C7

Fix the unit in Figure S-3: ng L^{-1}.

A: We fixed the units in the caption of Figure S-3 to be $ng\ L^{-1}$.

R1-C8

Fix the reference error in Table S-5 of the supplement.

A: We corrected the reference error in Table S-5 of the supplementary materials.

Response to Francesco Bregoli (Reviewer 3)

General comments

Methodology

R3-C1

The mechanism of water/soil partitioning (or absorption to soil) is complex. Here, the choice of their relative parameters is not process-based. It appears that are from back calculations or calibrations for the specific case of China (Grill et al. 2018) and may be not valid for other areas of the world. I understand that this is difficult for such global scale. But this needs to be better discussed.

A: We appreciate the thoughtful consideration of the complexities involved in the mechanism of water/soil partitioning and the associated parameter choices. We fully agree that the relative parameters used for the calculation of the direct discharge coefficient may not be universally applicable, as they were derived from a specific case study in China (Grill et al. 2018). While it would certainly be preferable to include region-specific parameter settings, the lack of data to support refinement of such parameters throughout the globe represents, at present, a considerable challenge. In recognition of this, we acknowledge that these uncertainties can substantially impact the outcomes, especially in regions with limited treatment infrastructure (see original manuscript lines 596-609). Nevertheless, we believe that the introduction of the direct discharge coefficient into the model is a crucial design feature and improvement of the model, since in previous large-scale contaminant fate models untreated pathways were not considered at all.

While recognizing the limitations, we would like to reiterate that our study is, in essence, a test of the model on a global scale. The limitations arising from unknown processes and measurement uncertainties are integral to this pioneering effort and set the stage for further refinement, development and future application of the model.

In order to further discuss this issue, as requested, and to reiterate the uncertainty while expressing the importance of the direct discharge coefficient, we added the following text after line 283:

"While these methods are simplistic in comparison to the real soil processes, no previous large-scale model considers untreated pathways as sources of contaminants, which can be substantial in regions with limited treatment infrastructure."

And we also revised and expanded the text at line 601:

"In fact, Grill et al. (2018) found in a sensitivity analysis for China that the setting of the direct discharge coefficient in rural areas represented the main source of model uncertainties. However, while the simplification of soil-related processes and the determination of their spatially heterogenous parameter settings remain a major challenge and likely source of error, in particular in areas dominated by untreated pathways, these simulations are critically important to be implemented in the model design. For example, in the presented case study, the untreated pathways contributed an estimated 72% of the global emission of sulfamethoxazole, demonstrating their decisive role. Despite the large uncertainties, the baseline scenario was able to reproduce field measurements reasonably well, especially considering the large range of possible values for the direct discharge coefficients (see Table 1). "

R3-C2

On-soil and groundwater contaminant degradation are here accounted with a degradation parameter based on the Euclidean distance from hypothetical source point and closer stream. But, groundwater flow does not necessary follow straight lines, but also follows flow directions trough positive gradients related to the local terrain geomorphology. Also, aquifers increase residence time, and therefore degradation. Because you state that the model is not very sensitive to this parameter, why did you add it on your model? This way, it seems that you add unnecessary complexity.

A: We agree that the hydrological flow path may be closer to reality than Euclidean distance in catchments that are governed by clear topography. But considering the uncertainty inherent in establishing hydrological flow paths at a global scale, including in challenging areas such as floodplains, we decided to simplify the process (it should be noted that in the original model implementation the flow path concept was used, see Grill et al. 2018). However, while the model was not 'very' sensitive to this process, we believe that accounting for the relative reduction in contaminant loading into rivers due to the location of population settlements and for variability in drainage density is an important model feature. For example, it may lead to improved simulations for particular regions which may not have been detected in the global sensitivity analysis. By including it in the model design, we hope it can be tested more rigorously in the future.

SMX case study and validation

R3-C3

SMX has both human and veterinary use. Therefore, it is complex to account only for human uses when validating the model. In the validation, you attempted to focus only on catchments were human use dominates. However, at global scale and with such big basins, this is very difficult.

A: We fully agree with this comment and acknowledge that to exclude other sources of antibiotics in measurements is virtually impossible, especially for large rivers. Our attempt was only to avoid measurements that explicitly reported in their description that other sources were included, to avoid known cases where measurements and predictions are by default expected to be incomparable. We also noted in the original manuscript that other sources are a possible cause of the negative bias observed in our model results, as discussed in lines 648 to 652. To clarify the methodology used, we changed the explanation regarding the selection of measurements in lines 370 to 374:

"In order to be selected for inclusion as a MEC in the model evaluation, the literature source must have reported the specific location (i.e., in the form of coordinates, river names, or river intersections) of the measurements. In addition, we discarded any MEC where the literature source explicitly mentioned that the dominant use of antibiotics in the catchment feeding the river was associated with veterinary or industrial activities, since the current version of HydroFATE is not adapted to account for these sources."

R3-C4

You defined several scenarios by changing relevant parameters. However, all parameters choice should be justified. For instance, you discussed the WWTPs removal variability in literature being 2% (min), 49% (ave), 73 (max) based on literature values. What about the other parameters in the different scenarios?

A: The excretion fractions and the instream decay constants are also based on literature sources, as described in detail in Section 4.1.1, as stated in the manuscript (line 392). The direct discharge coefficients of Scenarios 1 and 2 are taken from the existing study for China (Grill et al. 2018; see line 282)—we added this reference to line 392 (*"... and in Grill et al. (2018) for the direct discharge coefficients."*).

This leaves only the direct discharge coefficients for Scenarios 3 and 4 unexplained. Our goal with Scenarios 3 and 4 was to present minimum and maximum predicted concentrations within reasonable parameter variations. Since the direct discharge coefficient is a highly uncertain parameter developed based on a previous study (Grill et al. 2018), it has almost arbitrary variability. For this reason, we selected the values of 0.2 and 0.9 as plausible values that are close to the extremes of 0 and 1, which were analysed in the supplementary scenarios. To clarify this, we added an explanation to line 397:

"In the absence of relevant literature values, plausible boundaries for the direct discharge coefficients of Scenarios 3 and 4 were set slightly above 0 (with 0 representing complete decay along untreated pathways) and below 1 (representing no decay along untreated pathways)."

Finally, we rephrased the last statement in the Table caption to add clarity:

"For parameter settings and configurations see text."

R3-C5

High flow conditions, although favourable for higher dilution, is not considered as extreme low-end scenario.

A: As also indicated in the previous response above, Scenarios 3 and 4 are not supposed to represent "extreme" conditions, but rather reasonable/plausible conditions within the parameters' uncertainties. To avoid misunderstanding and to clarify that we are not using high-flow conditions, we changed the names of Scenarios 3 and 4 to *"Low-end case, average flow"* and *"High-end case, low flow"*.

R3-C6

The choice of using or not MECs below detection limit for validation is contradicted a couple of time in the MS. I would describe it better and univocally in the methodology section. In my opinion, MECs below detection limit are still important for model validation.

A: We apologize for not being clear when introducing the methodology associated with the selection of MECs. In the revised manuscript, we moved the following statement from line 558 to line 416:

"In addition to the 227 MECs, 134 measurements were classified as 'not detected' or 'not quantified.' To evaluate these cases, PECs at the same locations were verified to determine if they were correctly predicted to be below the detection or quantification limit (LOD or LOQ, respectively), depending on the limit reported by the study."

Furthermore, line 560 was corrected to:

"For PECs at the same locations as MECs that were reported to be below detection, 60% were correctly predicted to have concentrations that fall below the detection limit. If allowing an error of one order of magnitude, the success rate increased to 93%."

We also corrected the statement appearing at line 515:

"Modelling results were evaluated by comparing predicted SMX concentrations with available measurements in river reaches across the world using 227 MECs with values above the detection threshold and 134 measurements below detection."

R3-C7

You accounted the uncertainty in model prediction due to discharge condition into your PECs. If you include it again in MECs, it means that you are considering this uncertainty twice, which is not correct.

A: After due reflection and consideration, we respectfully disagree with this assertion. It is important to note that the uncertainties depicted in panel b of Figure 5 pertain to MEC uncertainties, rather than being indicative of model uncertainties. The purpose of including this graph was to illustrate how unknown characteristics specific to MECs might contribute to disparities between MECs and PECs at the same location. Such disparities should not be categorized as model uncertainties.

Nevertheless, in order to provide greater clarity on this matter, we have relocated panel b from Figure 5 (Section 4.4) to the Discussion section (Section 5.2), along with its accompanying explanation. This adjustment aims to facilitate a clearer differentiation between uncertainties related to model predictions and those associated with the evaluation data.

Discussion

R3-C8

I would appreciate a deeper discussion on the quality prediction (i.e. on NRMSE, NSE, PBIAS, KGE parameters). Are they expressing a good or bad prediction performance of your model? How do they compare with other similar large scale models performance? Is it an acceptable performance for predicting contaminants at global scale?

A: We agree that the goodness-of-fit indicators should be further discussed and, as such, we revised the text appearing at line 642:

"The results showed an overall reasonable predictive capability with the goodness-of-fit indicators NSE and KGE above 0.6, and with 79% of PECs being within one order of magnitude of reported MECs. This was despite the inherent uncertainties associated with assumptions made in the development of the model and those associated with estimates of the various model parameters and input datasets. Unfortunately, the lack of specificity of field measurements, for which literature sources generally do not provide enough information on the precise locations of measurements nor river discharge conditions, does not allow for a conclusive evaluation of the model under different modelling scenarios. It is noted that other global water quality models, which also simulate substance loads and concentrations, have reported similar values of NSE between 0.4 and 0.71 (Font et al., 2019; Harrison et al., 2019). However, a more detailed comparison between results from these models and HydroFATE is difficult as different substances and spatial resolutions were applied."

Extra comments extracted from the annotated pdf:

It seems that you are missing an important dataset of measurements, Wilkinson et al. (2020), https://doi.org/10.1073/pnas.2113947119. Above, you cited them but their database of MECs seems not in yours.

A: The dataset by Wilkinson et al. was only released in 2022, i.e., after our data collection and implementation phase had been completed. Adding these data to our assessment would represent a major amount of work (and thus substantial delay in publishing this work) as all measurements would need to be georeferenced to our river network in a quality-controlled way before being assessed. Hence, we refrained from utilizing these new data at this stage. However, we do understand the value of these additional data and plan to use them in the future.

I do not understand how width and discharge would help you in locate the MECs points.

A: The river network does not always match the exact location in reality, and the coordinates of MECs are not always precise. Therefore, river characteristics such as discharge and width can help differentiate between different rivers. For example, at a river confluence the MEC may be recorded to fall between the main river and a small tributary. If discharge or width of the measurement location were known, the correct river could be more easily determined.

Why not using high-flow to account for low-end case?

A: High flows as long-term monthly averages are not as meaningful for risk assessments. Nonetheless, we changed the scenario denomination and hope it is clearer now (see Response R3-C5).

References

Ehalt Macedo, H., Lehner, B., Nicell, J., Grill, G., Li, J., Limtong, A., and Shakya, R.: Distribution and characteristics of wastewater treatment plants within the global river network, Earth Syst. Sci. Data, 14, 559-577, doi: 10.5194/essd-14-559-2022, 2022.

Grill, G., Li, J., Khan, U., Zhong, Y., Lehner, B., Nicell, J., and Ariwi, J.: Estimating the eco-toxicological risk of estrogens in China's rivers using a high-resolution contaminant fate model, Water Research, 145, 707-720, doi: 10.1016/j.watres.2018.08.053, 2018.