Interactive comment on "Spatially-explicit estimates of global wastewater production, collection, treatment and re-use" *by* Edward R. Jones et al.

Anonymous Referee #1

The paper by Jones et al. provides a revised and consistent global outlook of the state of wastewater production, collection, treatment and re-use. It uses available country-level wastewater data and regression analyses to estimate information where it is unavailable. The year selected for the country-level data was 2015, and unavailable data were standardized to the same year using relationships with GDP. In addition, the authors downscaled the country-level data to a 5-min resolution grid using return flow data from the global water balance model PCR-GLOBWB. The downscaling was validated using European, US, and some global (yet less in numbers) records of wastewater treatment plants. Validation efforts delivered reasonable model performance indicators, and uncertainties were estimated using a bootstrapping technique.

The final data product provides a gridded map that includes quantities of wastewater production, collection, treatment and re-use at 5-min spatial resolution. Global water quality models and large-scale water assessments have been lacking this type of information in the past, so this paper is clearly a very important addition to the field. As this is a global effort, there are severe constraints regarding data availability and quality. The authors developed (and explained) reasonable approaches to overcome these problems. While some of their methods are based on speculation regarding the relevant processes, the ultimate test of such an approach is the validation of the results. I think the authors did a commendable job in their validation and comparisons, and I do appreciate that they reveal important shortcomings and clearly state that the results must be interpreted with caution.

So overall I think this paper presents an excellent global-scale effort to generate an advanced and novel gridded map of wastewater quantities. The manuscript is generally well written and very clearly structured. I strongly recommend its publication. I have a series of mostly minor comments that I list below. They are all written with the intention to further improve the manuscript. I also want to express my thanks to the authors for providing this important dataset to the research community!

Individual comments (line numbers refer to PDF version):

1. Title (and elsewhere): After reading the title ("Spatially-explicit . . .") my initial expectation was that the paper will describe explicit locations of wastewater production and collection (e.g. locations of wastewater treatment plants). Only when reading the manuscript, I realized that the dataset refers to a modeled distribution of wastewater quantities at sub-national scale. This is still great, but maybe a slightly different title could help avoiding this confusion, such as "Downscaled gridded model estimates of global wastewater. . ." or you could at least refer to "model estimates" rather than just "estimates".

2. There is no distinction made in the dataset between industrial and domestic wastew-

ater - which have very different characteristics and effects on environmental waters. It would thus be great to briefly discuss whether and how the combination of domestic and industrial wastewater may cause problems in the dataset, and in particular in the downscaling process. For example, industrial wastewater can be produced at locations with little correlation to population centers, i.e. at very distinct or remote locations compared to domestic wastewater - is this accounted for in the modeled return flows of PCR-GLOBWB? Also, I assume that the validation data cannot clearly distinguish between domestic and industrial wastewater as well? Related to this: Line 58 states for the first time that throughout the manuscript, domestic and industrial wastewater are not considered separately but lumped. As this is very important, it could be emphasized more, e.g. by referring more clearly to "combined domestic and industrial sources". This could also be done at other locations, where appropriate.

3. Lines 62 and following: You state that "wastewater treatment improves the quality of 'used' water resources" and this notion seems to prevail throughout the introduction and discussion. But while "wastewater treatment" as a process certainly has the GOAL to improve water quality, what about the fact that substances that are not or cannot be treated by treatment facilities can cause the opposite effect: in these cases, wastewater treatment plants can represent point sources of pollution, especially in the case of emerging contaminants. This has not been addressed in the paper and I thus encourage the authors to at least briefly reflect on the issue.

4. Table 1 (Standardisation to 2015): There is very little explanation in the text about the rationale behind the standardization methods. It seems the main assumption is a linear behavior of wastewater amounts based on GDP, right? This could briefly be mentioned in the text. Also, for collection and treatment: I do not understand why the values are divided by GDP per capita but then not multiplied by GDP per capita but by total GDP. Is this just a typo?

5. Figure 3: Very interesting figure. The one country that stands out to me as a surprise in wastewater production is Egypt. The Nile and Nile delta show also exceptionally high

values in Figures 4 and 6. There is no comment about Egypt in the manuscript. Any explanations on why Egypt has so high domestic and industrial wastewater amounts?

6. To my knowledge, the European dataset of wastewater treatment plants reports treatment capacity only in Population Equivalent (PE); however, the manuscript states that the volume flow rate was obtained based on a linear regression for plants reporting both parameters (PE and volume). It would be interesting to know how many plants included this information since the openly available dataset at the EEA website seems not to provide the volume flow rate.

7. Line 437 and following: You say: "This may occur due to discrepancies between the design (i.e. maximum) capacity of wastewater treatment plants . . ." Ok, that is a fair point. But the other option is that your model overpredicts in places without treatment plants and therefore underpredicts in places with treatment plants. This potential model bias should be acknowledged and briefly discussed.

8. Line 244-245: "For validating downscaled wastewater re-use, only plants (with treatment capacity > 1 million m3 yr-1) using tertiary or higher wastewater treatment technologies were considered." The rationale for this decision is not clear to me.

9. Figure 1 shows that there are data for only ∼half of global countries available (e.g. production data are available for 118 countries), yet 90% of population is covered by data. The only explanation for this is that all the high-population countries are included in those countries for which data exist, correct? It may be worth pointing this out as initially I was confused on how these numbers match up. I found some explanations later in the manuscript, but maybe this general fact could be stated earlier.

Minor comments (line numbers refer to PDF version):

Throughout the text, the authors use the expression "data" in singular form ("data is . . ."). I am more used to data in plural form ("data are . . .").

Also throughout the text, the authors use the expression "whilst" (many times). It

is my understanding that "whilst" may be perceived as 'archaic' in American English (https://en.wikipedia.org/wiki/While). So maybe use "while" instead?

Line 24 (and possibly elsewhere): The expression "significant" is often reserved for instances where it refers to statistical methods. Here, an alternative might be to use "substantial".

Line 25: replace "containing" with "comprising"?

Line 28: I suggest spelling out the first occurrence of SDG here (or remove the example)

Line 80: "that ensure" instead of "to ensure"?

Line 88: rephrase "whereby . . ." - maybe "which includes a target that . . ."

Line 89: add a space in "SDG 6.3"

Line 98: say "to a grid level of 5 arc-minute spatial resolution"

Figure 1: I cannot find the letters (a), (b) and (c) reflected in the figure, so it is difficult to find out which panels the caption refers to. Also, I suggest removing the asterisk from the figure and simply add the definition of 'population coverage' as part of the normal caption.

Line 117: add "(Table 1)" after ". . . databases"

Lines 125-126 (and elsewhere): I find the use the acronyms, the article "the", as well as the verbs not consistent here. I would say "GWI reports . . . whereas FAO reports . . . and UNSD reports" etc.

Line 126: say ". . . reported by UNSD"

Lines 139-140: repetitive use of "both" - delete one?

Table 1: start title with "Wastewater data sources and population coverage by region and economic aspects. . ." Also, the footnotes of the table could be shortened. E.g. the

square brackets always refer to the number of countries, so this could be explained once in the title and does not need to be repeated for each of the footnotes.

Line 166 (and possibly elsewhere): there are some instances where "per capita" is written as "per Capita" (capitalized)

Line 168-169: "Data was transformed, as appropriate, to ensure normality." This statement is not clear to me. Does it refer to what is called "sqrt" in Table 3, which I guess means that the square-root of values was calculated? Could both instances be clarified, e.g. by adding a little more information in this sentence here?

Table 2: spelling of "Agricultural Land" should be "Agricultural land"

Line 201: The eight regions are listed here for the first time, but no explanation is provided on what exactly defines these regions. Is this some official breakdown so that one could look up the countries that belong to each region?

Line 215: "occur" instead of "occurs"

Line 241: repetition of "both"

Line 243: "where aggregated PER CELL"?

Line 248: "were" instead of "was"

Line 254 and 257: While most explanations in this paragraph are in the past tense, two verbs are in present tense ("is expanded" and "is then compared"). Typo?

Line 261: "were" instead of "was"

Line 289: "Human Development Index" (as it was also capitalized elsewhere)

Line 290-291: say "was found to have the strongest influence on"

Line 311: "nations" (plural)

Figure 2: Could add the number of countries that are displayed in each panel, e.g. add

"n = ...". Also, the graphs may be clearer if using white instead of grey background.

Line 326: move "billion" after parentheses

Line 331: "being from" instead of "from"

Line 353: "World Health Organization's" (with apostrophe)

Line 362: "indicate" instead of "indicates"

Line 379: add commas before and after "respectively"

Line 391: delete "regions"

Line 397: "treat" instead of "treated"

Line 398: "than in" instead of "than by"

Table 4, title: Start with "Wastewater production, collection, treatment and re-use (billion m3 yr-1) by region and economic development level." I suggest changing "brackets" to "parentheses" (as brackets in American English would refer to squared brackets). And say "regressions" (plural).

Line 421: "yr-1" instead of "/year". Also, "occur" instead of "occurs"

Line 422: "collection . . . and treatment . . . are" (instead of "is")

Line 426: "with only available wastewater resources"âĂŤnot clear, wrong wording?

Line 456: I suggest using "underpinning source data" instead of "underlying source data"

Line 472: "upon which country-level estimates incorporate" seems to be incorrect wording

Line 479: I suggest breaking this very long paragraph into 2 here

Line 486: say "factors" instead of "drivers" to avoid repetition

Line 489: say "of untreated" (add "of")

Line 496: "Whilst our results also rely on this approach, we instead used. . ." This sounds odd (first you say you do the same, then you say 'instead') - rephrase?

Line 512: The expression "acreage" may not be known to all readers. Could just say "spatial extent"

Line 519: close the parentheses

Line 522: "for as a baseline for"? should this be "as a baseline for"?

Line 525: ". . . problems of discrepancies in data reporting years and missing data are overcome." It sounds quite optimistic that the problems are truly "overcome" (i.e. solved). Maybe say "reduced" instead?

Lines 528-529: repetitive use of "particularly"

Lines 549-551: you use the word "such" three times here

Lines 550-551: The grammar/verb of the description of point (4) seems incorrect. Change "creating" to "create"?