

Review ESSD-2021-214 WWTP

High quality thoughtful effort to produce HydroWASTE.

Review link works, data download and open easily.

Do the authors correctly describe a database, or more properly a dataset. From looking at other recent examples in ESSD, this reviewer understands a database as open, accessible, recruiting (sometimes in easy automated fashion) additional incoming data (often from individuals as data providers), searchable through a variety of database fields, adopting and promoting novel community definitions, etc. Here a reader finds a static product (will the post-review version be on-line in interactive form?), with inputs entirely and solely at the discretion of the authors via 'official' national reports. Nothing in text about adding new data? Very good product, but probably more a dataset than a database? Produced with persistent skill but - even compared to HydroATLAS - this product seems more limited? Perhaps necessarily given the topic but authors provide no justification?

Lines 54, 55: exposure to chemicals from dense populations "which at a regional level can help prevent negative effects and determine hotspots of contamination." This in a discussion of importance of dilution factors but "prevent negative effects" requires higher dilution factors while "determine hotspots" implies lower dilution factors. Perhaps authors mean 'or' rather than 'and'?

Line 126: "basis for calculating the contaminant loads" but authors already informed readers that for many so-called contaminants of emerging concern, WWTP accumulate but do not ameliorate contaminants of concern. Level of treatment irrelevant in that case? Authors could write here 'loads of treated or untreated contaminants'?

Line 335: "underestimations" of population served, of WWTP spatial density, of? A reader needs clarification here? In Fig 3, underestimation refers to population served, e.g. number of WWTP must be underestimated in view of population served? Where are the missing WWTP? Not addressed in uncertainty discussion, e.g. "reflecting the incompleteness of WWTP records" but no attribution noted or solutions proposed?k Authors tend to consider weaknesses in data sources (e.g. OSM) or in European methods for estimating populations served, but fundamental question of missing (or absent) WWTP not addressed? If comparison data products (e.g. Jones 2021, and note they describe their product as 'results' or as a 'study') come from same sources and therefore propagate same weaknesses, authors have no alternate forms of validation? Given variations among countries in terminology and reporting, basing comparisons on only South Africa seems very weak?

Line 363: here reader finds that "10,445" (20% of total WWTP in HydroWASTE) WWTP discharged into large (lake or ocean) bodies of water. But, back at line 326, readers learned that 224 (of 58,500 WWTP in HydroWASTE) discharged into oceans or lakes. Why this apparent discrepancy. If this reviewer missed the reason, other readers will likewise miss?

Line 706, Figure 6: Here authors show waste water fractions whereas throughout most of the manuscript they a focused on dilution ratios. If the latter represent the most-preferred regulatory factor, why not adopt them consistently. Or, if national reports differ, explain the difference? Not clear to this reader why we confront both DF and % WW fractions?

Line 752, Figure 3: Reference at top of column 7 should be Jones et al. 2021? Authors of this manuscript have taken global treated wastewater total from Jones et al. of 188.1×10^9 , converted to daily average to get 514? (I get 515?)