

Geosci. Model Dev. Discuss., referee comment RC3
<https://doi.org/10.5194/gmd-2022-147-RC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-147

Christina Quaassdorff (Referee)

Referee comment on "Yeti 1.0: a generalized framework for constructing bottom-up emission inventories from traffic sources at road-link resolutions" by Edward C. Chan et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-147-RC3>, 2022

Thank you very much for this interesting study. The manuscript presents a bottom-up modeling framework based on HBEFA emission factors to estimate emissions at the city level. In this case, as an example, the framework is applied to the city of Berlin in order to assess the emission results in comparison to previously published data for the city.

General comments to the text:

- The level of detail that the framework needs for the input information and the resolution it can achieve are not clearly defined.
- Currently the text reads very specific on file formats and model running processes. Suggestions to focus on the application of the framework and demonstrate the applicability are made.
- Some additional analyses are suggested to demonstrate the accuracy of the framework for all the modelled pollutants.

Main comments to the text:

The study claims that high temporal and spatial resolutions are needed (Page 1 line 24-27). Please specify what is meant by high spatial and temporal resolution in this study and justify how this aligns with existing literature (potentially with microscale emission modelling). For the spatial resolution, it would help to add the average length of the roads and the minimum and maximum lengths used in this study. For the temporal resolution, are the 1-hour intervals enough, have those been aggregated from more disaggregated

information? Is that enough to capture traffic flow fluctuations that influence emission levels? Please specify and introduce in the text.

A clear scope of the proposed framework is needed. In the current manuscript, the Yeti framework seems to be mostly based on the use of HBEFA (an already existing model) so that the novelty and/or additional knowledge added is not clear to me. A clear description of the added value of Yeti would help to understand better the utility, capabilities, interactions with existing knowledge and potential limitations of the proposed framework.

Specific comments:

- Page 2 line 29-30: Please specify which databases are meant when referring to “integration of vehicle-level emission factors estimated from existing databases”.
- Page 2 line 49-51: Are COPERT, MOVES and HBEFA, considered in this study as high resolution emission models?
- Page 2 line 54: Please be specific on the details and resolution the bottom-up approach can provide. In this case, specifying the resolution that can be achieved and the specific details that can be provided would help to understand better the capabilities and requirements of the bottom-up approach.
- Page 3 line 65-72: Is Yeti a traffic emission inventory or a tool to calculate traffic emission inventories based on existing models and data? Please revise the text accordingly, in the current version it is not clear to me. Also, how flexible is the Yeti approach with the input information? Could it use second-by-second information and also with yearly averages? What are the limitations related to input information resolution. Since this framework seems to aim to work with varying levels of detail, did you analyze how the framework performed when using different levels of resolution for the input data?
- Page 3 line 71: Please explain the concept of “process symmetry”, is not interpretable from the current text.
- Page 3 line 77: What is meant by vehicle sectors and which specific ones have been investigated? Stating at this point a clear scope of the study would help to understand better the proposed framework capabilities and potential limitations.
- Page 3 line 76-80: Please clearly state the goals. The current text seems more as a description of what is done but makes not clear to me what are the goals to answer.
- Page 3 line 83: how long are the road segments in average? And what are the maximum and minimum length those could have for the proposed framework to estimate emission accurately?
- Page 3 line 84: where is the geometrical attribute of grade for each road segment coming from? Please specify in the text.
- Page 3 line 84: please provide a definition for road segment, and road level. Also, along the text road link is mentioned. Please unify and/or provide definitions.
- Page 3 line 88: Please provide a definition for “vehicle subsegments”.
- Page 5 line 120: Please provide a justification on why selecting only the temperature is the better option instead of selecting any of the other two options. Have you perform an analysis of the variables that have a stronger effect on cold start emissions or is this

based on literature? If the first, please explain. If the second, please add the corresponding references. What is the associated error for the cold exhaust emissions calculation attributed to this selection?

- Page 5 line 123-125: How are the cold start counts inferred from the hourly traffic count and the road type so that cold start events are identified? Please add explanation in the text.
- Page 5 line 126-127: Why is the dimensionless factor set to 0.3? Please explain in the text the rationale behind selecting this specific number.
- Page 5 line 135-136: Please be specific with the terminology, I guess you are referring to emission factors from evaporative emissions
- Page 5 line 143-144: what is the share of fuel-injected vehicles Yeti assumes in the vehicle fleet? Is that a case-specific share? Please specify in the text.
- Page 6 line 149: please be specific. Despite this framework seems to be mostly based on HBEFA, this manuscript should read as a standalone text containing all the essential information to follow the story. In this specific case, how many are the engine stops considered in HBEFA or how are those classified, and how does this translate to this specific study? Please, revise the manuscript for similar expressions.
- Page 6 line 151-152: What is referred to by direct data? Did you perform measurements? Please explain. Also, what percent of engine stops are coming, in this study, from direct data and what percent is estimated? Did you perform a comparison on how well the proposed estimation works compared to the "direct data" of number of engine stops?
- For each of the emission processes (non-exhaust, evaporative, cold and hot exhaust), please specify which pollutants are estimated in Yeti.
- Page 6 line 170: It seems that for all the other emission processes HBEFA 3.3 was used and only for non-exhaust PM HBEFA 4.1 was used. What are the implications of using two versions of the model to obtain different sets of emission factors for the different emission processes? Could the framework be adapted to be updated with current and coming versions of HBEFA to calculate emission inventories with the most updated emission factors available? The last is partially raised in the Summary at the end of the manuscript, but it seems that a specific update would be required with each update of HBEFA. An estimation on how fast Yeti can adapt to updates of the emission model would be useful.
- Page 7 line 177: The term road links is used here. Different terminology is used throughout the text, previously road level, and road segments were mentioned. Please introduce a definition for each of them in the text.
- Section 2.2 seems to be very specific on the modules used, but I would be interested in knowing how those influence the emission calculation. For example, it is mentioned in lines 178-179 that Yeti can operate on a subset of the traffic network. Does that also imply that if for a specific subset very detailed traffic activity input information is available more detailed emission outcomes could be obtained? Can Yeti consider that in some way? How small can the network subset be? And on the other hand what is the bigger network that would still be an option to run with Yeti. For the last, it would help if the authors provide a processing time and data size quantification for this specific study in Berlin.
- Section 2.2.1 Data organization, section 2.2.2 User-specified configuration and section 2.2.3 Execution flow: these sections are lacking the specific relationship to the study done in Berlin. It kind of reads as a user manual of Yeti. For example in line 195 it would help to know the specific execution parameters that were used in this study if relevant to explain some of the outcomes. Please revise, I think this is a good opportunity to show the applicability of Yeti and not place the focus only on how the framework internally works.
- Page 7 line 190: several new concepts are introduced in this line (emission strategy, day type, meteorological profile), please provide a definition for them.
- Page 9 lines 241-243 in reference to Table 2: some symbols from table 2 need additional explanation such as IDTS.

- Page 9 Section 3.1.1. HBEFA emission factors and field data: please clearly state which ones are the field data specified in the title of this section and how those were obtained and used in Yeti. A note on how that could be replicated in other cities and which would be the needed information to obtain would be helpful.
- Page 9 section 3.1.3. This section has a lot of details on how the files look like and the directories, but is lacking an explanation of the content of table 3. This also applies to previous section 3.1.1. Please, it would help to add an interpretation of the information presented in the tables to properly understand them. If not relevant for to explain the results in the main paper, maybe this part could be moved to the supplementary material.
- Page 10 line 268: Why would you have road segments with zero length? And what is the error associated to ignore the segments that have no indicated traffic direction? Are those a significant amount of your segments?
- Page 10 line 274: please be specific with the highest resolution that can be achieved. How long are the individual road segments, and what would be the highest temporal resolution?
- Page 10 line 280: please add the essential information to understand how the data recompilation was done. What period of time was covered with the measurements? What kind of roads were measured? How big was the sample size obtained and how were those postprocessed to be obtain a full dataset to use in Yeti?
- Page 10 293-294: Why was finally this attribution selected and how is this attribution of LOS IV vs LOS V affecting the results for this specific example?
- Page 12 line 356: results for the other pollutants (CO and PM) would be also interesting to see in the supplementary material. That would help to understand the applicability of the proposed framework, explain relationships between pollutants and accuracy of the outcomes for all the modelled pollutants in this study.
- Page 13 section 4.1 (line 365): In figure 4, NO_x emissions for summer and winter are the same. Can Yeti consider differences in traffic volumes and traffic composition for different seasons or even lower temporal resolutions (month, week, days,...)?
- Page 13 Section 4.2. related to the NO_x results, was fuel type also considered to analyze the outcomes? what is the average vehicle fuel distribution (% of diesel, gasoline,...) in the example roads for the different vehicle types (passenger cars, bus, motorcycle...)? Together with current tables A1 and A2, this additional table would help to understand better the outcomes presented in tables 10-11.
- Page 15 Summary section: please revise to add notes on the scope and potential limitations of the framework raised previously.
- Figure 5: add to the caption the average length of the segments.

Other minor comments:

Page 3, line 63: Revise English in the phrase “but the cost and the level configurability play a significant role”

Page 3, line 66: revise the expression “road network traffic conditions and emission factors are large”. This sentence is not clear to me.

Page 4 line 109: typo, "LODs" seems it should be "LOS"

Page 8 line 226: "from the" is repeated in "from the from the City of Berlin (Diegmann et al, 2020)"

Page 12 line 342: Revise the sentence "an increase in aggregate evaporative increase in evaporative diurnal HC emissions..."

Page 13 line 383 and line 398: busses

Page 19 line 518: $\delta \square \square \frac{1}{2}$ is included in table 1 not table 2.