

EDITORIAL

Reviews and responses for

Compilation of an open-source traffic and CO₂ emissions dataset for commercial aviation

Authors: Antoine Salgas, Junzi Sun, Scott Delbecq, Thomas Planès, and Gilles Lafforgue

Reviewers: Antonio Franco and Max Li

Editor: Martin Strohmeier

1. Original paper

DOI for the original paper: <https://doi.org/10.59490/joas.2023.7201>

2. Review - round 1

2.1 Reviewer 1

In this paper, authors present a methodology to aggregate different sources of flight information in order to obtain an open-source air traffic dataset and they apply it to obtain one for 2019. They also consider an existing aircraft performance surrogate model to obtain the CO₂ emissions corresponding to that air traffic dataset. As an additional contribution, the authors have developed and made available a tool to ensure user-friendly access and exploration of the open-source dataset.

The references cited are pertinent, thorough enough, and up to date. The methodology is sound and appropriately described. The section addressing the validation of results through external databases includes interesting and sensible reflections and explanations for global, regional and country-wise discrepancies.

Notwithstanding the above, some improvements could be made:

1. Authors are suggested to indicate the percentage of traffic-undetermined routes among the total routes and the percentage of data whose aircraft type information was unknown. This way the readers would have a clearer awareness of the estimations being performed by the authors.
2. An accuracy metric for the fuel consumption regression model has not been provided, unlike what happened with the seat capacity regression. Furthermore, although the predicted curve in Figure 5 (page 9) fits reasonably well with original data for larger distances (let's say above 4000 km), it is not the case for shorter distances. In fact, the original data seems to show a bimodal behaviour for shorter distances. This fact has an important negative impact on prediction accuracy for short-haul and medium-haul flights. This reduced accuracy may be behind the fact that larger errors are obtained for CO₂ emissions of domestic flights in Table 4 (page 12). Would it be feasible for authors to explore, in this current manuscript, a different regression strategy for flights with a distance less than 4000 km, maybe considering clustering to account for data multimodality? (See comment in page 14, line 408.)

3. When estimating the number of seats through XGBoost (Section 2.3.3, page 7), the authors are strongly encouraged to provide not only accurate results but also estimates of feature importance. Apart from providing valuable insight, this would also help to identify candidate features to be replaced (see comment on page 14, line 404).

Additional minor comments:

- a. References are not ordered the same way as they are cited; reordering is strongly encouraged.
- b. There is a portion of Somalia (specifically, Somaliland, an unrecognized country in the North-West) missing in Figure 6. Even if there is no data assigned to that region, authors are recommended to depict the coastline.

All in all, my recommendation is to accept the paper subject to minor revisions (those needed to address the previous comments).

2.2 Reviewer 2

The authors present a rigorous, meaningfully comprehensive procedure for assembling worldwide aviation flows through the combination of different open-source data sets, and apply their compiled data set to the purpose of CO2 emissions estimation of worldwide commercial air traffic flows. This paper should be a canonical example of "doing the most with what we have" in terms of open-source aviation data sets. Not only is the method transparently described (both in terms of drawbacks—which there are many, although this is understandable given the constraint of only using publicly available data—and advantages), but the authors have also created a well-documented Github repository for others to use their worldwide estimated air traffic flow data for other purposes outside of sustainability benchmarking and emissions estimation. I only have one comment regarding the topic of data inequity across worldwide geographies, but this was a very strong and interesting paper to read from my perspective.

I very much commend the authors on Figure 6, and would suggest that a more in-depth discussion be given regarding what this reveals about the relative inequities in data access/availability, and how this has downstream impacts on creating sustainability goals (such as those considered in this paper), but also for ATM global harmonization efforts in general. How are such harmonization efforts (e.g., multi-regional TBO, etc.) supposed to be effective if we don't even have basic demand/schedule data uniformly across different regions? The countries where there is a severe under- or over-estimation in ASKs reveal locations where the current "standard" data sources do not provide much insight. I don't think I have come across another analysis that puts this issue in such stark focus: I recognize that this is not the main purpose of the paper (it is focused on sustainability/emissions estimates), but a small paragraph on this data inequity would be highly appreciated.

3. Response - round 1

3.1 Response to reviewer 1

The authors would like to thank you for the comments that helped me prove the paper. The answers to the various remarks are given as follows.

1 - Authors are suggested to indicate the percentage of traffic-undetermined routes among the total routes and the percentage of data whose aircraft type information was unknown. This way the readers would have a clearer awareness of the estimations being performed by the authors.

Response

Thanks for the suggestion. Some information has been added to address it, at two different steps of the process. First, among the Wikipedia parsed routes, 41% of the routes remain traffic-undetermined after merging all the sources. Then, the share of each source for the seats offered in the compiled database has been added to Table 2, as well as the ASK. We have also added a sentence to specify the share of ASK for which the aircraft is unknown, although the value can now be directly read from the table.

2 - An accuracy metric for the fuel consumption regression model has not been provided, unlike what happened with the seat capacity regression. Furthermore, although the predicted curve in Figure 5 (page 9) fits reasonably well to original data for larger distances (let's say above 4000 km), it is not the case for shorter distances. In fact, the original data seems to show a bimodal behaviour for shorter distances. This fact has an important negative impact on prediction accuracy for short-haul and medium-haul flights. This reduced accuracy might be behind the fact that larger errors are obtained for CO₂ emissions of domestic flights in Table 4 (page 12). Would it be feasible for authors to explore, in this current manuscript, a different regression strategy for flights with a distance less than 4000 km, maybe considering clustering to account for data multimodality? (See comment in page 14, line 408.)

Response

Thanks for the remark. Indeed a strong bimodal behaviour appears in Fig. 5 but did not catch our attention first. Further investigation revealed that this behaviour was found only on US-BTS sourced data as can be seen in the figure below, on which the source of each flight was coloured. Given that each source used has mostly the same aircraft types, the greater dispersion of BTS points are linked to the higher accuracy of the seating capacity estimate. Indeed, BTS reports for each [origin-destination-airline-aircraft] type tuple the number of seats available. This information is not available when using radar-based data such as OpenSky or Eurocontrol and an average fleet value seating capacity is used for each aircraft type instead, as mentioned in the article. When the regressions are done separately for each source, they almost coincide as highlighted by the lines on the plot below.

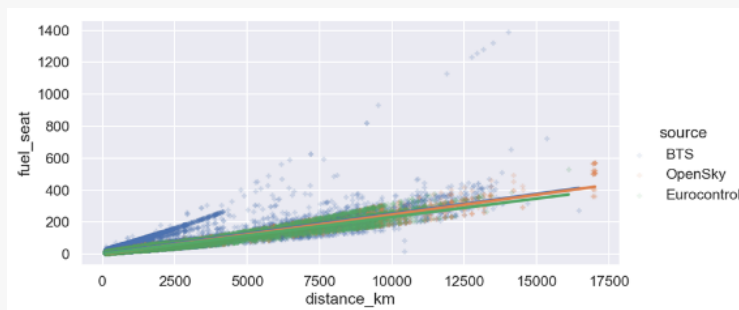


Figure 1. Source-specific regression, analogous to article Fig.5

Indeed, most of the highly dispersed points were found to be related to particular use cases. We found in the raw data that the clear trendline appearing below 4000 km was for instance a fleet of Delta Air Lines Boeing 757 dedicated for charter flights to transport among other NBA teams, with a 72-seat configuration, compared to the usual 160 to 220 seats of the aircraft type. There are other specific all-business routes such as “La Compagnie” service from Paris to New York. These flights correspond to a much higher fuel burn per seat than usual. However, these flights are a minority, as illustrated by the modified scatterplot given below (Fig.2), on which the size of each data point was made proportional to

the number of flights represented. Although these specific flights are a minority, this investigation led us to discover a weakness in the regression. It was performed unweighted, meaning that all data points contributed equally to the regression, no matter the number of flights represented. Considering the weights in the regression reduces drastically the importance of those business-charter flights. However, as illustrated by the comparison of unweighted and weighted regression in Fig.2, both regressions are very similar.

The bimodal behaviour that could be seen in the original figure is less important once points are weighted. Hence we do not consider that the clustering approach suggested would improve drastically the results and it would be hard to affect this "charter" cluster performance on the aircraft- undetermined routes. Therefore, we suggest using the weighted regression that seems to capture well the average fuel burn per seat trend on the whole distance range. As a reminder, this regression is used only when the fuel burn cannot be computed directly. We added the r^2 as suggested. In the modified manuscript figure, we propose to only display the weighted regression that inherently represents air traffic better. As for your remark on larger errors obtained on domestic flights, I would rather attribute them directly to the existing surrogate model used (and on which the regression is performed): even in aircraft-information-rich regions, the error on CO₂ is greater than on ASK (see Europe and North America in Table 4). Investigating this and modifying the aircraft performance model is beyond our objectives for this paper but would indeed be a very nice improvement.

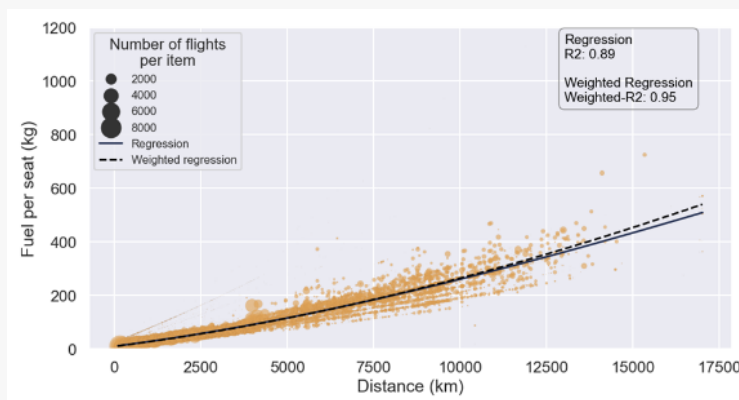


Figure 2. Weighted regression, improvement to article Fig.5

3 - When estimating the number of seats through XGBoost (Section 2.3.3, page 7), the authors are strongly encouraged to provide not only accuracy results but also estimates of feature importance. Apart from providing valuable insight, this would also help to identify candidate features to be replaced (see comment on page 14, line 404)

Response

Thanks for this relevant remark, considering that much could still be done in my opinion to improve the model. We've added a feature importance figure as an appendix, as was the case for the presentation, accompanied by a comment.

4 - References are not ordered the same way as they are cited; a reordering is strongly encouraged.

Response

For us the references appear the same way they are cited, it's just that Figure 1 is a tikz figure and references are cited inside the figure. Therefore, these references are accounted for in the first citation of the figure.

5 - There is a portion of Somalia (specifically, Somaliland, an unrecognized country in the North-West) missing in Figure 6. Even if there is no data assigned to that region, authors are recommended to depict the coastline.

Response

Indeed, there is no data for Somaliland, everything is affected to Somalia in the airport database used. For some territories, a mapping was made to affect the emissions to the main state (like for French Overseas departments and territories, which have a country code different than the one of France, or Greenland/Denmark). I propose a modified version of the figure that superposes a grey layer for these countries.

3.2 Response to reviewer 2

The authors would like to thank you for the comments that helped improve the paper. The answers to the various remarks are given as follows.

Comment: I very much commend the authors on Figure 6, and would suggest that a more in-depth discussion be given regarding what this reveals the relative inequities in data access/availability, and how this has downstream impacts on creating sustainability goals (such as those considered in this paper), but also for ATM global harmonization efforts in general. How are such harmonization efforts (e.g., multiregional TBO, etc.) supposed to be effective if we don't even have basic demand/schedule data uniformly across different regions? The countries where there is a severe under- or over-estimation in ASKs reveal locations where the current "standard" data sources do not provide much insights on. I don't think I have come across another analysis that puts this issue in such stark focus: I recognize that this is not the main purpose of the paper (it is focused on sustainability/emissions estimates), but a small paragraph on this data inequity would be highly appreciated

Response

Thanks for your advice, I think that it is indeed worth mentioning that in the manuscript. I've added a paragraph to discuss this issue. The source fusion process is complex and time-consuming, and increases the risk of Merging increases the risk of data errors. It also leads to unequal treatment of data depending on the source, which is a handicap when it comes to comparing countries rigorously. Ideally, all authorities should make these very aggregated data available, as the Bureau of Transportation Statistics does. This would make no difference to the privacy of flights, which are already accessible on a case-by-case basis to everyone via FlightRadar24 or any flight comparator. For now, this compiled dataset suits our needs of a baseline for regionalised prospective scenarios, but we're not yet ready to assess the decarbonization strategies of small aviation markets for instance