

EDITORIAL

Reviews and Responses for **Compilation and applications of an open-source dataset on global air traffic flows and carbon emissions**

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

Reviewers: Luis Delgado, Tamara Pejovic, Duc-Thinh Pham

Editor: Michael Schultz

1. Original paper

DOI for the original paper: <https://doi.org/10.59490/joas.2024.7365>

2. Review - round 1

2.1 Reviewer 1

The paper uses open-source datasets to estimate the movements and supply of aviation worldwide.

The paper suggests a refreshing approach which is very interesting and could lead to many different applications. This second part is somehow one of the drawbacks of the article. I feel that the authors present too many different applications, which can distract the reader from the main message and outcome. Some of the material could be moved to the Appendices or even kept for another research.

The methodology to generate the routes, flight and passenger demand is interesting and novel.

When the authors describe how data was collected from Wikipedia, they should explain if some effort has been made to validate the data contained. It's not clear what they mean by "Others may be missing from the list of airports used in the first place...", please clarify that. The authors refer to Wikidata, is that the same as Wikipedia?

It would be useful to provide a list of all the features used to estimate the traffic. The authors mention that some parameters, such as the GDP per capita and Gini coefficient of incomes, are added to the list of features, but there's not a comprehensive list of the features used. Figure 17, in the Annex, provides the importance of the features, but these are not described. If the focus of the paper is to describe the methodology and approach to generating the data, the information on the features and analysis of their importance should be considered part of the main text of the article.

For the traffic estimation, the authors mention using average aircraft capacities and load factors. From where are these obtained? They also identify issues with the data sources' reliability, e.g., mislabeling flight origin-destination from OpenSky. How did the authors deal with these issues?

The authors mention that the first 1000-quantile was used to remove data with missing features. What do they mean by that?

In Section 2.4, the authors identify the priority order of the datasets, but they refer to them just by

their reference (22, 23, 24, 26, 26, 27). It would also be useful to mention them by their name so that the reader has a clear understanding of the datasets used and their prioritization without having to go to the reference section.

For the estimation of the emissions, the authors could also consider the work from Montlaur *et al.*, Analytical Models for CO₂ Emissions and Travel Time for Short-to-Medium-Haul Flights Considering Available Seats, Sustainability, which provide analytic models for short and medium distance flights based on the available seats per aircraft.

In the validation, when the authors mention that with respect to IATA, the open-source dataset has a difference of around 1%, do they mean IATA with respect to the outcome of their modelling?

In Figure 6, are the errors assigned to the country of departure?

The tool for exploring the dataset is useful. However, as mentioned, the examples of applications on hydrogen-powered aircraft and transport inequalities are interesting, but I think they diminish the value of the article. They are very well elaborated, but therefore, quite long overall; they require additional datasets and modelling assumptions. My advice would be to focus on the first part of the article, expand the analysis on the features and modelling assumptions and quality (validation) of the estimation of demand and emissions, provide these as examples in an Annex, and just summarise the main findings as examples of applications that can be done with the dataset. These, with some expansion, could be their own research, and even if thoroughly analyzed, they provide just some examples of the range of potential analyses that could be done. Other aspects, such as how the transport data can be updated and maintained over time, could be of higher interest to the reader.

There is a typo in the open data statement where it says “Not that external inputs” when it should be “Note that external inputs”.

2.2 Reviewer 2

The paper presents a comprehensive methodology for creating a global air traffic and CO₂ emissions dataset based entirely on open-source data, which is a commendable effort towards enhancing the transparency and reproducibility of aviation-related environmental research.

The use of innovative data sources, such as Wikipedia, to obtain airline route information showcases a novel approach to filling data gaps left by traditional sources. Furthermore, the development of the AeroSCOPE tool significantly enhances the accessibility and usability of the dataset for a wide range of potential applications.

However, there are areas where further improvement could enhance the overall impact and reliability of the study. The following are some constructive criticisms and suggestions for future work.

1. Methodological Considerations

1.1 Data completeness and accuracy: While the paper acknowledges the variable accuracy of the dataset at different aggregation levels, it could benefit from a more detailed discussion of the implications of these inaccuracies for potential users. A sensitivity analysis could be valuable in understanding how variations in data quality across regions or routes might affect research outcomes or policy recommendations derived from the dataset.

1.2 Estimation Model validation: The paper provides an evaluation of the estimation model’s performance but could be strengthened by including a more thorough validation against independent datasets that were not utilized in the model development. This external validation could provide further confidence in the model’s generalisability and accuracy.

1.3 Handling of Intra-Country inequalities: The research notes the limitations of the dataset in capturing within-country inequalities. Future work could explore methodologies for integrating more granular socio-economic data to model these intra-country variations more effectively. This could involve collaboration with national statistical agencies or leveraging emerging data sources such as social media or mobile phone data to infer patterns of air travel usage among different population segments.

2. Environmental Impact Assessment

2.1 Life Cycle Emissions of Alternative Fuels: The paper discusses the potential of electric and hydrogen-powered aircraft in reducing CO₂ emissions but does not fully account for the life cycle emissions associated with these alternative energy sources. Future extensions of this work could incorporate life cycle assessment (LCA) methodologies to provide a more holistic evaluation of the environmental benefits and trade-offs associated with transitioning to these new technologies.

2.2 Mitigation strategy comparisons: The dataset offers a valuable tool for assessing the impact of various decarbonization strategies on the aviation sector. Future studies could use the dataset to conduct comparative analyses of different mitigation measures, such as fuel efficiency improvements, operational changes, market-based measures, and the introduction of low-carbon fuels.

3. Technological and Policy Implications

3.1 Infrastructure and Regulatory Considerations: The analysis of electric and hydrogen aircraft's network potential is insightful but would benefit from further exploration of the infrastructure and regulatory changes needed to support these technologies. Future research could investigate the logistical, technological, and economic challenges associated with deploying alternative fuel infrastructure at airports and the regulatory frameworks required to facilitate the adoption of these new aircraft types.

3.2 Broader environmental and social impacts: While the focus on CO₂ emissions is critical, a more comprehensive assessment of the environmental and social impacts of aviation—including air quality, noise pollution, and economic effects on communities—would provide a more rounded view of the trade-offs involved in different aviation decarbonization pathways.

2.3 Reviewer 3

This is an interesting and excellent work. It proposes an approach for creating an open-source air traffic dataset, which was utilized to gather data for the year 2019. The framework aggregates various sources of flight information, addresses missing flight data, and ultimately provides estimates for CO₂ emissions. The accuracy of this open-source dataset was evaluated against four reference datasets. Furthermore, the authors have developed an interactive tool, AeroSCOPE, to facilitate dataset exploration and have discussed two applications. The manuscript is well-written, and the case studies are thoroughly examined. To provide further clarity, my comments are outlined as follows:

1. It is somewhat confusing regarding the main contributions or objectives of this work, as they are delineated differently in the title, abstract, introduction, and conclusion sections. Although this work is part of the AeroMAPS development, there isn't a sufficient scientific justification for its significance.

2. Is it accurate to characterize this work as an extended version of a conference paper [20], with the addition of Section 4 on applications and use cases? Apart from Section 4, I have not observed any other extensions of this work beyond the conference paper [20].

3. If I understand correctly, the feature list developed in Section 2.3.2 will be utilized to estimate the number of seats on each route. It would be beneficial to include a table showcasing this list.

4. Regarding seat capacity estimation, has there been any methodology discussed in the literature? Random Forest and XGBoost appear to yield promising results, with an R^2 of approximately 0.72. How can we justify that this performance is sufficient for subsequent steps, such as CO2 estimation?

5. Regarding figures:

- A coverage map of the open-source dataset should be included to provide an overview of the currently available dataset, which would be useful for comparison with Figure 6.

- The authors should reconsider the captions of Figures 8 and 9 to better reflect the content and scope of the illustrated information.

The continents are referenced differently in Tables 3 and 4. There are still some spelling and grammatical errors that should be double-checked and corrected.

3. Response - round 1

3.1 Response to reviewer 1

The paper suggests a refreshing approach which is very interesting and could lead to many different applications. This second part is somehow one of the drawbacks of the article. I feel that the authors present too many different applications, which can distract the reader from the main message and outcome. Some of the material could be moved to the Appendices or even kept for another research.

Response

Thanks for the review. Indeed, the core of the research is the dataset presented and evaluated in the first three sections of the paper. As mentioned in the introduction, this submission to JOAS is an extended version of a paper presented at the OpenSky symposium in 2023, and it is specifically completed by adding the application section. The possibility of extending the conference papers was suggested by the organizers, and we thought it was a great opportunity to provide a starting point for further research by ourselves or the open-source community. Moving the application to the appendix would make the paper equal to the conference paper; we chose, therefore, to keep them as they were. However, a sentence has been added in the introduction to highlight that the core of the research is the compilation of the dataset. Similarly, a sentence has been added at the beginning of the application section to emphasize the prospective dimension of these applications.

When the authors describe how data was collected from Wikipedia, they should explain if some effort has been made to validate the data contained. It's not clear what they mean by "Others may be missing from the list of airports used in the first place...", please clarify that. The authors refer to Wikidata, is that the same as Wikipedia?

Response

This remark is relatively similar to one of Tamara Pejovic's comments. Some insights on the limitations and implications of using Wikipedia as a source to fill the gap are given the related answer. To complete the answer by points more specific to your comment, there is no systematic validation of the existence of a route in the parsing process, due to the difficulty of collecting the sources quoted on each airport page. The most important limitation is, as outlined in the article, that the data is collected as it is on the Wikipedia page, at the time of the collection. In our case, data was collected in 2023. Numerous routes have been disrupted or created since 2019. One could expect that this is the main driver for the

differences instead of wrong sources, but there is no evidence to support this affirmation. Finally, the validation is made at the aggregated level by comparing the resulting dataset to a commercial reference.

Concerning the "Others may be missing from the list of airports used in the first place...", the sentence was clarified in the text and replaced by "Another issue are airports that would not be listed in the original set of airports [35] that is parsed to collect the associated destinations. To address this problem, any airport identified during the destination parsing process that is not already in the original set of URLs is subsequently added to this set. Therefore, the only way an airport can be missed is if it is neither listed among the Wikipedia airport pages nor served by any flight originating from an airport in the original list." It seems that this better describes the 2 step process used. First, destinations from all "known" airports are parsed. If it reveals "unknown" airports in those destinations, those are also explored to collect their destinations.

Concerning Wikidata, a sentence was added to define Wikidata, a structured data structure behind Wikipedia and other projects of Wikimedia Commons.

It would be useful to provide a list of all the features used to estimate the traffic. The authors mention that some parameters, such as the GDP per capita and Gini coefficient of incomes, are added to the list of features, but there's not a comprehensive list of the features used. Figure 17, in the Annex, provides the importance of the features, but these are not described. If the focus of the paper is to describe the methodology and approach to generating the data, the information on the features and analysis of their importance should be considered part of the main text of the article.

Response

All the features added to the route dataset are mentioned in Section 2.3.2, although not in a formal list. Such a list has been included as a section outline. The relative feature importance is now discussed in the main text at the end of Section 2.3.3.

For the traffic estimation, the authors mention using average aircraft capacities and load factors. From where are these obtained? They also identify issues with the data sources' reliability, e.g., mislabeling flight origin-destination from OpenSky. How did the authors deal with these issues?

Response

The average aircraft capacities are obtained from the Planespotters fleet database, as mentioned in Section 2.2. A clear reference to this has been added. Regarding radar data source completeness, the main strategy to mitigate the impact of data quality is to use this source as a last resort. Due to coverage similarity, in particular with Eurocontrol and BTS datasets, the source remains very marginally used for model training, being used to determine the traffic on less than 2% of the model training dataset. Sentences to explain this were added in Section 2.3.3.

The authors mention that the first 1000-quantile was used to remove data with missing features. What do they mean by that?

Response

The 1000-quantile was evoked as an imputation strategy for missing features of some data entries. A sentence was added to define the 1000-quantile (value so that 99.9% of the values taken by the data are higher). The 1000-quantile was used with the idea that missing features are linked to a "small" airport or country. This imputation choice is arbitrary; this also justifies the choice to use a regressor able to handle missing values instead.

In Section 2.4, the authors identify the priority order of the datasets, but they refer to them just by their reference (22, 23, 24, 26, 26, 27). It would also be useful to mention them by their name so that the reader has a clear understanding of the datasets used and their prioritization without having to go to the reference section.

Response

Thank you for this wise remark; this has been corrected in the two passages concerned.

For the estimation of the emissions, the authors could also consider the work from Montlaur *et al.*, Analytical Models for CO₂ Emissions and Travel Time for Short-to-Medium-Haul Flights Considering Available Seats, Sustainability, which provide analytic models for short and medium distance flights based on the available seats per aircraft.

Response

We have added a reference to this paper, as well as another model, as a way to provide insights for a reader who would like to try alternative fuel burn models.

In the validation, when the authors mention that with respect to IATA, the open-source dataset has a difference of around 1%, do they mean IATA with respect to the outcome of their modelling?

Response

The passage in question was clarified. There are 1% more available seats worldwide in 2019 in this work than reported by IATA.

In Figure 6, are the errors assigned to the country of departure?

Response

Yes, these are errors on the total departure ASK for each country. A clarification was made.

The tool for exploring the dataset is useful. However, as mentioned, the examples of applications on hydrogen-powered aircraft and transport inequalities are interesting, but I think they diminish the value of the article. They are very well elaborated, but therefore, quite long overall; they require additional datasets and modelling assumptions. My advice would be to focus on the first part of the article, expand the analysis on the features and modelling assumptions and quality (validation) of the estimation of demand and emissions and, provide these as examples in an Annex and just summarise the main findings as examples of applications that can be done with the dataset. These, with some expansion, could be their own research, and even if thoroughly analyzed, they provide just some examples of the range of potential analyses that could be done. Other aspects, such as how the transport data can be updated and maintained over time, could be of higher interest to the reader.

Response

As explained in the first response, the structure of the article makes the application essential as it is the only difference from the conference paper that is being extended. We understand, however, that this could diminish the importance of the dataset, but we think that providing insights on how this data could be leveraged to decarbonize air transport remains an interesting outcome of the article. Moreover, the first application is a direct consequence of the data through an open-access visualization website.

We have added a sentence in the conclusion to take into account your remark on data maintenance over time.

There is a typo in the open data statement where it says “Not that external inputs” when it should be “Note that external inputs”.

Response

Thanks for the remark.

3.2 Response to reviewer 2

Suggestions - Methodological Considerations

- Data completeness and accuracy: While the paper acknowledges the variable accuracy of the dataset at different aggregation levels, it could benefit from a more detailed discussion on the implications of these inaccuracies for potential users. A sensitivity analysis could be valuable in understanding how variations in data quality across regions or routes might affect research outcomes or policy recommendations derived from the dataset.

Response

Indeed, the accuracy of the dataset is variable but sufficient for the intended primary use case. For clarification and to avoid misuse of the dataset, a paragraph was added at the end of Section 3. In particular, we recommend users perform a sensitivity study with uncertainties representative of the error levels reported in the related tables and figures. In the primary use case, i.e. calibration of partitioned decarbonization scenarios, such analyses will be made in future works.

Estimation Model validation: The paper provides an evaluation of the estimation model’s performance but could be strengthened by including a more thorough validation against independent datasets that were not utilized in the model development. This external validation could provide further confidence in the model’s generalisability and accuracy.

Response

Four external comparison sources are used in Section 3. Country-level energy consumption was compared against IEA-reported kerosene consumption. Total ASK were compared to IATA figures. The most important comparison sources are a commercial dataset of OAG (however, only the year 2018 was available to us and was converted using a uniform growth rate to “2019” values) and the compiled, country-level data of ICCT, whose work is based on the OAG dataset. None of these sources were used in the training process. Commercial sources such as Cirium or Sabre would be a nice extra backup but are out of financial reach so far.

Handling of Intra-Country inequalities: The research notes the limitations of the dataset in capturing within-country inequalities. Future work could explore methodologies for integrating more granular socio-economic data to model these intra-country variations more effectively. This could involve collaboration with national statistical agencies or leveraging emerging data sources such as social media or mobile phone data to infer patterns of air travel usage among different population segments.

Response

This dataset has no insight at all on who takes air transport. There are few studies on the subject, and particularly the one used as a reference in the inequalities plot. Another example of an interesting reference (<https://doi.org/10.3917/socio.102.0131>) looks at the air transport democratization in France between 1978 and 2008. A sentence has been added in the inequalities section to acknowledge these potential prospects.

Suggestions - Environmental Impact Assessment

Life Cycle Emissions of Alternative Fuels: The paper discusses the potential of electric and hydrogen-powered aircraft in reducing CO₂ emissions but does not fully account for the life cycle emissions associated with these alternative energy sources. Future extensions of this work could incorporate life cycle assessment (LCA) methodologies to provide a more holistic evaluation of the environmental benefits and trade-offs associated with transitioning to these new technologies.

Response

Indeed, such work is being done in another work package of the AeroMAPS project. The idea here was just to assess the market potential in terms of "emissions on routes that could be operated with technology X" and not of "emissions that would be abated with technology X". We have added further clarifications in the dedicated section to acknowledge this.

Mitigation strategy comparisons: The dataset offers a valuable tool for assessing the impact of various decarbonization strategies on the aviation sector. Future studies could use the dataset to conduct comparative analyses of different mitigation measures, such as fuel efficiency improvements, operational changes, market-based measures, and the introduction of low-carbon fuels.

Response

As mentioned in the introduction and the conclusion, this work is part of a more global project, AeroMAPS. This platform is made to perform such analyses, and the dataset presented here serves as a calibration database for the base year of a prospective scenario on which AeroMAPS builds its analyses.

Infrastructure and Regulatory Considerations: The analysis of electric and hydrogen aircraft's network potential is insightful but would benefit from further exploration of the infrastructure and regulatory changes needed to support these technologies. Future research could investigate the logistical, technological, and economic challenges associated with deploying alternative fuel infrastructure at airports and the regulatory frameworks required to facilitate the adoption of these new aircraft types.

Response

Similarly, some of the very relevant questions raised here will be answered by use cases of AeroMAPS or some other prospective simulation platforms. However, in section 4.2, the approach consisting of selecting n airports to maximize the market covered is part of the discussion: with a given budget to install hydrogen infrastructure, where should we start? Or conversely, to reduce by $x\%$ the air transport emissions, which airports should be equipped and at what cost.

Broader environmental and social impacts: While the focus on CO₂ emissions is critical, a more comprehensive assessment of the environmental and social impacts of aviation—including air quality, noise pollution, and economic effects on communities—would provide a more rounded view of the trade-offs involved in different aviation decarbonization pathways.

Response

As highlighted in aviation LCA studies (for instance: <https://doi.org/10.2514/6.2022-1028.c1>), most of the environmental of current air transport comes from the CO₂ emissions and fossil resources depletion. However, the adoption of biofuels, e-fuels, or batteries necessitates more resources of various natures while being less carbon-intensive, which justifies a comprehensive LCA approach. This is planned for further developments.

Clarification Needed: The use of Wikipedia to fill in missing flight information is innovative but raises questions regarding data reliability and consistency. Wikipedia content can be edited by anyone, potentially leading to inaccuracies.

Suggested clarification: "The choice to utilize Wikipedia as a primary source for filling data gaps is predicated on the comprehensiveness of its airline and airport listings. However, we acknowledge the dynamic nature of this data, given Wikipedia's open editing model. To mitigate potential inaccuracies, we employed a systematic validation process, cross-referencing information with authoritative aviation databases where available. Further details on the validation steps, including the handling of discrepancies between Wikipedia and official sources, are essential to assure the scientific community of the data's reliability."

Response

Although Wikipedia is indeed a community source, it requires authors to cite as many sources as possible. Theoretically, it is possible to collect the corresponding sources during the parsing process, but that was not made to ease the data collection and also to make it more robust (A mandatory 2-column Airline-Destination table is in each Wikipedia airport page, but references are either in a third column, or in plain text, or next to each destination). The probability of persons voluntarily altering an airport page is not null indeed but is considered non-systemic per hypothesis (i.e. if a particular airport network is not very well estimated, the consequences overall are limited). When the compiled dataset is evaluated against a commercial dataset (see Figure 1), two types of errors are tracked: either a "delta", meaning the route is both in Wikipedia and OAG but the traffic is not the same, or an "extra"; i.e. routes in one or another dataset. Although there are some "extra" for the "estimation" dataset (i.e., based on Wikipedia parsing), it remains reasonable. The figure is not included in the main article (it is in the supplementary notebooks), but this is mentioned in the text in section 3: *"Two types of error can be distinguished at this global level. The first case is when the route is recorded by both datasets, but with different volumes: around 387 Mn seats are "lost" in this case. The second case is when routes are in either dataset but not in the other: 280 Mn seats are on routes not referenced in OAG data, and 122 Mn in the opposite case."* While estimation represents 29% of the total seat capacity in the compiled dataset, it contributes to 40% of this "extra" type error. The most limiting concern about using Wikipedia comes from the inability to track the date at which the network collected is valid. In our case, as Wikipedia was parsed in April 2023, the corresponding network is certainly more representative of 2022 than 2019. Section 2.3.1 was modified according to your suggestion and to emphasize the restricted reliability of using Wikipedia.

Concern: The paper mentions estimating CO₂ emissions using an existing aircraft performance surrogate model without detailing the assumptions, limitations, or the model's applicability across diverse aircraft types and operational conditions. It's imperative to elaborate on the model's underlying assumptions, its validation against real-world emissions data, and how it accounts for variations in aircraft operations. Additionally, discussing the model's limitations and potential impact on our dataset's overall emissions estimations will enhance methodological transparency.

Suggested revision: "To estimate CO₂ emissions accurately, we utilized a surrogate model based on aircraft performance. This model's effectiveness varies across different aircraft types, operational

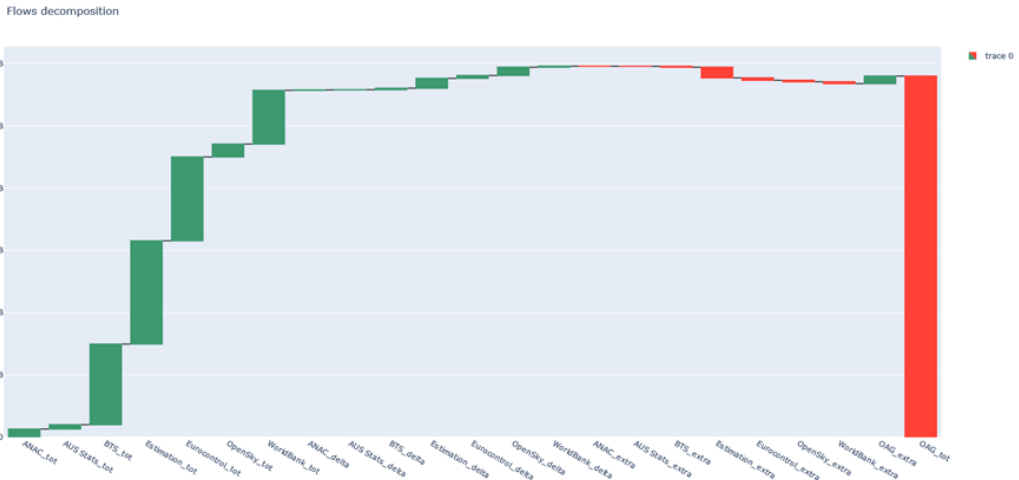


Figure 1. Error decomposition compared to OAG dataset.

profiles, and environmental conditions. ...”

Response

Thanks for the remark. The model was only briefly evoked. A dedicated paragraph was added in Section 2.5 to describe more carefully the model used and to give more information about the reported accuracy.

Clarification Needed: The paper briefly describes combining various open-source datasets to achieve global coverage but lacks detail on the specific methodologies employed to ensure consistency and accuracy after aggregation. Clarifying the harmonization techniques, such as data normalization procedures, handling of overlapping data entries, and resolution of conflicting information, will provide deeper insights into the robustness of the aggregated dataset.

Suggested Revision: "The methodology for aggregating diverse open-source datasets into a coherent global database is critical for ensuring dataset integrity. The process involved aligning disparate datasets on a common set of parameters, including flight routes, aircraft types, and traffic volumes."
"

Response

Certainly, the methodology combines 6 different datasets and uses Wikipedia-based estimation to bridge the remaining gaps. Combining these many sources is necessary to extend the data coverage. To avoid interference, only one source is used on each route, and a priority order, based on the relative quality, is established in the aggregation process. This step is now described in more detail in Section 2.4. Despite the precautions, large differences between the sources remain because of their nature. An extra warning on this point is added at the end of the section.

The use of regression and machine learning models to estimate missing flight information and traffic volumes is mentioned, but the paper could benefit from a more in-depth discussion on model selection, training data set preparation, validation processes, and the handling of missing or incomplete data.

Suggested Enhancement: Add further detail on the selection criteria for these models, the prepro-

cessing steps undertaken to prepare the training dataset, and the techniques used to validate model predictions against known data points. Additionally, revealing the strategies for dealing with missing or incomplete data within the training process would bolster the credibility and applicability of the estimation models employed.

Response

The regression models are used only to estimate the traffic volume on each route found by scrapping Wikipedia. Missing flight information (aircraft type, airline, etc.) is not inferred, although that would be a nice prospect. The detailed notebook, including preprocessing, is given in the GitHub repository mentioned in the supplementary material. A sentence was added to define the 80/20 train-test split principle and, thus, the technique used to validate the model against known data points.

The preprocessing required by the linear regression has been further detailed. The log-linear regression paragraph was extended for better understanding. A comment on the interest of those first two techniques for higher-level prediction (such as market forecasting) was added as a nuance to their poor performance on this specific route-level prediction.

The concept of a regression tree was not defined. A sentence was added specifically, and the link with the random forest regressor is defined with more details. Both the random forest and XGBoost algorithms are also defined with more caution.

Finally, a sentence was added at the end of Section 2.3.3 to emphasize the fact that there might be other regression possibilities.

Suggestions:

1. Original: "This paper presents a method to obtain an open-source air traffic dataset for 2019, first introducing a method to aggregate different sources of flight information, then identifying and completing missing flight information using Wikipedia parsing, and finally estimating CO2 emissions using an aircraft performance model."

- Suggestion for revision: "This study develops an innovative open-source dataset detailing 2019's global air traffic flows and associated CO2 emissions. We outline a comprehensive approach that combines diverse flight data sources, addresses data gaps through systematic Wikipedia parsing, and implements an aircraft performance model to estimate CO2 emissions. Methodology promises reinforced reproducibility and broader accessibility in aviation environmental research."

Response

Thanks for the suggestion; we have added it to the paper without using the first person, however.

2. Original: "The overall process for obtaining an open-source database is described in this section. 2019 has been chosen as the reference year for building this database, and the following years have been disrupted by COVID-19."

- Revised: "The methodology for compiling an open-source database is comprehensively detailed in this section, with 2019 selected as the baseline year due to the significant disruptions in air traffic patterns caused by the COVID-19 pandemic in subsequent years."

Response

Thanks that is included "as is" in the revised version.

3. Original: "For evaluating the climate impact of aviation emission scenarios, including alternative fuel aircraft, the analysis does not fully consider the life cycle emissions."

- Revised: "To comprehensively evaluate aviation emission scenarios' climate impact, including the introduction of aircraft powered by alternative fuels, it's imperative to expand our analysis to encompass life cycle emissions. This holistic approach will ensure a more accurate assessment of the environmental benefits and potential trade-offs."

Response

We could not find this passage in the text submitted. A clarification made following one of your previous comments seems to answer this remark.

4. Original: "While the accuracy of the dataset varies for different routes, major traffic flows at the country and continental levels are reasonably well estimated."
- Revised: "Accuracy levels within the dataset exhibit variation across different routes. However, it's important to note that estimations of major traffic flows are robust at both country and continental scales, albeit with room for refinement to ensure uniform data reliability."

Response

It seems that this remark applies to a similar sentence in the abstract that was modified accordingly.

5. Original: "The research notes the limitations of the dataset in capturing within-country inequalities in air transport usage."
- Revised: "This study acknowledges the dataset's current limitations in accurately reflecting the disparities in air transport usage within individual countries, underscoring a crucial area for future methodological enhancements."

Response

A sentence to highlight this research prospect was added in the conclusion.

6. Original: "This work provides a customizable dataset for aviation environmental research, paving the way for a deeper understanding and monitoring of the sector's carbon footprint."
- Revised: "This manuscript introduces a pivotal resource in aviation environmental research—a customizable dataset that empowers researchers to delve into comprehensive analyses of the aviation sector's carbon footprint. Future enhancements and broader applications of this dataset stand to significantly advance our collective efforts towards sustainable aviation."

Response

Thanks! This has been added at the beginning of the conclusion, although with a more neutral tone.

3.3 Response to reviewer 3

It is somewhat confusing regarding the main contributions or objectives of this work, as they are delineated differently in the title, abstract, introduction, and conclusion sections. Although this work is part of the AeroMAPS development, there isn't a sufficient scientific justification for its significance.

Response

Thank you for this comment, which helps us to take a step back from how we highlight our objectives. We have modified the abstract to better highlight the objectives and contributions of the paper: the compilation of an open-source dataset, as well as research avenues for its exploitation. A more detailed outline of the work is also given in the introduction, and the focus is recentered around the paper instead of AeroMAPS at the beginning of the conclusion.

Is it accurate to characterize this work as an extended version of a conference paper [20], with the addition of Section 4 on applications and use cases? Apart from Section 4, I have not observed any other extensions of this work beyond the conference paper [20].

Response

As you mentioned, the paper is similar to [20] for the first three sections, albeit with some modifications due to the different reviews. The possibility of extending the conference papers was suggested by the organizers, and we thought it was a great opportunity to provide a starting point for further research by ourselves or the open-source community. A concern on the significance of Section 4 was also pointed out by Luis Delgado, but we think that identifying ways of exploiting the data collected is a valuable addition to the article. Some revisions have been made in the introduction to better highlight the structure of the article, with parts 2 and 3 being more central to the work.

If I understand correctly, the feature list developed in Section 2.3.2 will be utilized to estimate the number of seats on each route. It would be beneficial to include a table showcasing this list.

Response

As the features were somehow hidden by the surrounding text, this point is appropriate. We've added a proper list of the features collected at the end of the section.

Regarding seat capacity estimation, has there been any methodology discussed in the literature? Random Forest and XGBoost appear to yield promising results, with an R^2 of approximately 0.72. How can we justify that this performance is sufficient for subsequent steps, such as CO₂ estimation?

Response

Most of the literature focuses on gravity models, such as the one tested in the paper for demand forecasting. We have added two references providing extended reviews. In our case, the gravity model estimate is very low, most probably because of the compiled nature of the training data. Although this is an innovation, it comes with limitations: for instance, the granularity of the data is variable, with many small charter flights in some regions (BTS/Eurocontrol data). While this makes data richer (these flights happened), it also creates some pollution for the regression. An alternative for having better results with those standard models would be to focus on only one kind of source, thus losing the global prediction capacity of the model. Also, the way our features are collected, with web parsing, makes the presence of missing values inevitable, making gravity models use problematic. Sentences to clarify this point were added to section 2.3.3. The flexible and robust nature of regression trees is more suited to this context. The R^2 of 0.72 seems sufficient to do estimates at the regional or country level, given the validation conducted in Section 3. In particular, the estimation of both traffic and CO₂ emissions appears unbiased. However, as we try to highlight it, the CO₂ estimates and comparisons of individual routes are not within the objectives of the work. Higher fidelity would be needed in this case. Advanced techniques, such as neural networks, are also presented in the literature but have not been tested.

A coverage map of the open-source dataset should be included to provide an overview of the cur-

rently available dataset, which would be useful for comparison with Figure 6. The authors should reconsider the captions of Figures 8 and 9 to better reflect the content and scope of the illustrated information.

Response

This comment is very interesting because the resulting map is rich in information. It is difficult to map the coverage of the existing open-source datasets because several sources coexist for many countries. Instead, we propose to add a map showing the fraction of ASK that is estimated for each country in the main text (see Figure 2, corresponding to new Figure 7 on the paper). Countries with a lot of estimates seem to be well estimated, in particular, thanks to the recalibration step of the flows on each estimated route with respect to the residual traffic at each airport. On the other hand, the large errors found in many low-traffic countries have more to do with a discrepancy between the OAG reference and the World Bank dataset used for compilation. We have added this point to the analyses.

Captions of figures were modified based on the scope.

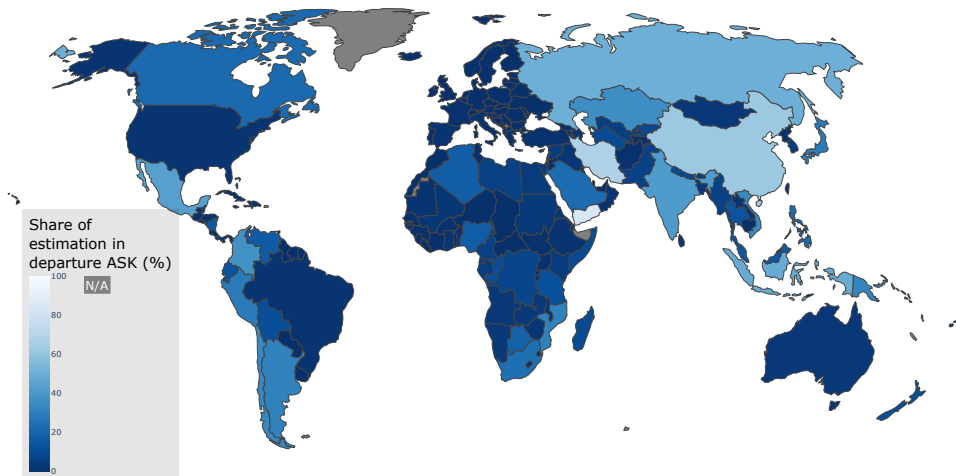


Figure 2. Share of estimation as a source in total departing ASK.

The continents are referenced differently in Tables 3 and 4.

Response

Table 4 notation was extended to Table 3, as it avoids using unnecessary acronyms.