

EDITORIAL

Reviews and Responses for Automating the Estimation of Noise and Emissions near Airports with ADS-B Data

Authors: Gonçalo Soares Roque and Johannes Reichmuth

Reviewers: Anandini Jayanthi, Céline Parzani, and Saskia Drossaart van Dusseldorp

Editor: Manuel Waltert

1. Original paper

DOI for the original paper: <https://doi.org/10.59490/joas.2024.7901>

2. Review - round 1

2.1 Reviewer 1

The aim of this paper is first to develop an automated process for the estimation of environmental impacts for historical scenarios, specifically noise and pollutant emissions in the vicinity of airports, by using open-source data such as ADS-B data. In a second step, this automated process is applied to the Cologne Bonn Airport, open-source data for the year 2019 are then coupled with confidential ones (airport flight logs and noise measurements) from the Cologne Bonn Airport in order to perform a validation especially for noise estimation.

The topic is interesting. Nonetheless, there are aspects that could be refined to improve the overall impact and reliability of the study. My questions and comments are the following:

- Scope and motivation should be more clearly described. Line 30, 31 need to be more detailed. It is not so clear what exactly this new framework is helping us with?
- The paper's originality in relation to existing works needs to be further elaborated.
- The introduction should be restructured to outline the existing literature and clarify the position of the paper regarding it. For instance, some previous works such as Sarrat, C.; Aubry, S.; Chaboud, T.; Lac, C. Modelling Airport Pollutants Dispersion at High Resolution. *Aerospace*2017, 4, 46. <https://doi.org/10.3390/aerospace4030046> should be mentioned.
- It would be interesting to mention some details about GRAPE library and to recall the main assumptions of the implemented models.
- Concerning meteorological data, how can the choice of the IEM specific weather database can be justified? (why not ERA5 data for instance)? what is the resolution for meteorological data? Moreover, is it possible to precise the way "the closest point to the trajectory" is defined (line 142, 143)
- Which environmental impacts are exactly considered? The expression "local air quality" appears several times in the paper but dispersion of pollutants does not seem to be considered, so this point

has to be clarified.

- In what sense the first library of the automated process is an extension of "traffic" library?
- Assumptions have to be mentioned clearly, for instance "always using the maximum available thrust for all departures" only appears in Section 4.3.
- The Section 3.1 may be improved given a title to each step (filtering, smoothing...). The part from line 157 to 186 is very detailed compared to the others of the automated process. If it represents a main contribution of the paper, this point has to be mentioned early. If not, the content may be synthesized.
- Concerning noise and emission results sections, it would be helpful, additionally, to compare the results with those from other studies or models.
- In section "Noise results", in Table 2, how the discrepancy on the results for the station M06 between arrival and departure can be explained?
- What are the assumptions mentioned line 360? It is possible to perform a test to rule out one of the 3 causes proposed for discrepancies? The corresponding part describing the differences observed in Figure 4 and 5 has to be reformulated in order to better emphasize the selected cause.
- At last, how the emission results obtained with the automated process are in agreement with the ones from others studies, potentially on others airports?

Minor comments:

- Citations: Pretto *et al.* line 23, it would be better to give the reference in this way Pretto *et al.* in [6], [7] have focused...
- In citation [14], authors are not mentioned.
- A reference is missing for Doc 29 line 24.
- The sentence line 42 could be reformulated.
- Give the details of the acronym FOI line 65 and a reference for this database.

2.2 Reviewer 2

this was a well-written paper, congratulations. I would like to suggest the following changes:

1. Line 166, a citation is required for taking 90% MLW.
2. In Equation 1, γ is the descent angle, not the climb angle.
3. Figure 2 should include the runways more clearly or are to be described.
4. Line 347 establishes threshold for selecting ADS-B noise event by considering absolute LAMAX values of recording stations. These thresholds are selected without comparing and correlating the validation set and ADS-B set. The threshold selected could very well be too high for arrivals, hence the under-prediction. It would be a good idea to include it in this paper, or put it down as something to improve in the future.

2.3 Reviewer 3

I would like to commend the authors for their innovative idea and approach. Estimating noise and emissions using open-source data represents a valuable contribution to the field. This methodology offers significant benefits. It is much more (cost-)effective than extensive in-situ campaigns, which the authors have implicitly acknowledged. Additionally, while airport operators often monitor this

data, they are sometimes hesitant to share it. The use of open-source data bypasses this issue, enhancing accessibility and transparency. Furthermore, the development of a reliable tool that can be applied to any airfield and provides uniform output is particularly valuable. Such an approach enables intercomparison studies that are rarely achievable with airport-provided data, as airports may have limited interest in supporting such analyses.

Major remarks:

The paper lacks a clearly stated aim. A model framework has been developed and applied to a case study, but a clear research question is missing. This absence results in a conclusion that feels more like a summary, leaving the reader uncertain about whether the presented tool performed well in the case study. A clearly defined research question would benefit both the reader and the authors.

While the title suggests a focus on the estimation of noise and emissions, the paper's primary emphasis seems to be on the automation, modeling, and evaluation of arrival and departure trajectories at an airport using publicly available ADS-B data. The calculation of noise and emissions appears to be presented as an application of this approach. This impression is based on the following observations:

Methodological focus on trajectory modeling: While the calculation and processing of the ADS-B data are described in detail, the GRAPE software itself is described only briefly, with limited information provided, unless the reader is familiar with various existing guidelines. A more detailed description is not available elsewhere, such as on the associated GitHub repository, leaving the impression that studying the source code is necessary to understand the underlying methodology. Additionally, the discussion on limitations focuses solely on issues related to the trajectories and thrust settings and neglects the assumptions underlying the calculations performed by GRAPE, such as the LTO cycle. Although these assumptions adhere to international standards, their limitations play a crucial role in assessing the effectiveness and reliability of the noise and emissions estimation method. A more comprehensive discussion of these factors would greatly enhance the estimation of the methodology's overall accuracy and potential impact.

Lack of comparison to existing practice(s) of noise modelling: While the methodology for computing noise data and its comparison to actual measurements is reasonably well-detailed, it is unclear whether the authors' approach improves upon or is at least comparable to established methods, such as those described in ICAO DOC 9911 or comparable studies the authors have considered in the introduction (sources 3 and 4). Including a comparison with existing methodologies would add valuable context for evaluating the effectiveness and relevance of the approach developed by the authors. This would help clarify its role within the state of the art for aircraft noise estimation.

Lack of context and critical discussion of emissions: I believe the authors' approach has significant potential to improve emissions estimations based on the LTO cycle. Incorporating real trajectories and weather conditions, especially for thrust computation, presents a notable step forward. However, some significant shortcomings remain with the chosen approach. For example, while weather conditions are partially accounted for in thrust calculations, their broader impact on emissions and their impact is not addressed. The primary focus of the discussion is on comparing calculated values for time spent in arrival (LTO: approach) and departure (LTO: takeoff and climb) modes. An additional difference lies in the thrust settings, which are approximated in the authors' method but fixed in the LTO cycle. While these differences highlight deviations in real trajectories worth observing, the resulting variations in emissions are a logical consequence of these adjustments and, as such, not inherently significant. Essentially, the authors state that "we change variables that impact emissions, and we observe changes in emissions," without providing any context as to whether these results are meaningful or valuable. A comparison with in-situ measurements, as was done with noise, other established approaches to emission modeling (e.g., the methods outlined in ICAO DOC 9889), or even a discussion of the broader implications of these findings, would add much-needed

context and elevate the discussion of emissions in the paper.

In summary, the reconstruction of airport traffic using ADS-B data is an interesting approach with promising implications for the remote estimation of noise and engine emissions near airports. However, assessing this potential requires a more comprehensive comparison with state-of-the-art methods and a stronger, evidence-based discussion of the findings. For the presented study, a thorough and critical discussion of its limitations would add valuable insight.

Minor remarks:

It is not entirely clear what kind of weather data has been used, in what way. The text discusses “airport and time specific weather data” (line 59), “Airport weather observations” (Table 1), or “the weather report closest to each trajectory point” (line 142) while the schematic in figure 1 contains METAR. I assume this means that METAR reports were used throughout. Is this also used to calculate true air speed for the trajectory segments near 3000ft altitude?

Setting the taxi time to zero really provides data only “near” airports and not “at” airports. While the authors never claim the latter and discuss arrival and departure throughout the paper, I believe it would be helpful to explicitly mention that traffic at the airport itself is not considered in this estimation.

It is not explained and how the threshold levels for noise described in line 312 and following are chosen. Where do these values come from and why is there no difference between day and night for two stations?

The flight results show that approximately 81.5% of the number of flights were identified by the developed method. It is not clear to me if and how these roughly 20% difference are considered when comparing the modelled noise events with the measured ones.

3. Response - round 1

We would like to thank the reviewers for their comments and insights. The paper has been updated and the response to each comment is provided below.

3.1 Response to reviewer 1

Scope and motivation should be more clearly described. Line 30, 31 need to be more detailed. It is not so clear what exactly this new framework is helping us with?

Response

We have completely rephrased the second part of the introduction to more clearly state the purpose and objectives of this work, as well as its contribution and scope (lines 29-48).

The paper’s originality in relation to existing works needs to be further elaborated.

Response

See response to previous comment. The contribution of the paper in relation to other work has been put more in focus.

The introduction should be restructured to outline the existing literature and clarify the position of the paper regarding it. For instance, some previous works such as Sarrat, C.; Aubry, S.; Chaboud, T.; Lac, C. Modelling Airport Pollutants Dispersion at High Resolution. *Aerospace* 2017, 4, 46. <https://doi.org/10.3390/aerospace4030046> should be mentioned.

Response

The introduction has been restructured to more clearly state the contribution of this work. Regarding the position of the paper to existing literature, we believe the first part of the introduction now addresses the most recent publications in the context of using ADS-B data for environmental impact calculation purposes. Regarding the suggested paper for mention, it focus on modelling high-resolution emissions dispersion. The emissions dispersion process is highly dependent on localized meteorological conditions and is outside our scope, this has been further clarified in the restructured introduction (lines 46-48).

It would be interesting to mention some details about GRAPE library and to recall the main assumptions of the implemented models.

Response

The GRAPE tool abstains from making any assumptions regarding its implemented models. On the contrary, it aims at providing a one-to-one implementation of environmental impact calculation models, and provides the user with as much control as possible regarding both the input data used as well as any control parameters the models may accept. This is presented to the reader in Section 3.2 (lines 221-223). Furthermore, this section also details for the automated process the models, input data and control parameters used for estimating both fuel flow, noise, gas pollutant emissions and particulate matter emissions. Describing the assumptions inherently made by each model (i.e. not in the user's control) is in our opinion outside the scope of this work, as the models used are well established standards.

Concerning meteorological data, how can the choice of the IEM specific weather database can be justified? (why not ERA5 data for instance) ? what is the resolution for meteorological data? Moreover, is it possible to precise the way "the closest point to the trajectory" is defined (line 142, 143)

Response

The introduction now clearly states that emissions dispersion is outside the scope of the paper. As high resolution meteorological data is not required by noise, fuel consumption or emissions inventories models (e.g. BFFM2) at an airport scenario level, ERA5 data (or other 3D grid weather data) is not applicable. The paper uses IEM as a source of openly available METAR data, which is more accurate than using ISA standard values or airport year averages. Regarding the last point, the wording has been reformulated to make it clear that for each trajectory point of an operation (arrival or departure) occurring at a given airport, the respective METAR report closest in time to the trajectory point timestamp is used (lines 167-169).

Which environmental impacts are exactly considered? The expression "local air quality" appears several times in the paper but dispersion of pollutants does not seem to be considered, so this point has to be clarified.

Response

The scope of this work has been better clarified in the restructured introduction. It is now clearly stated directly in the introduction which pollutants are considered, and that emissions dispersion is outside the scope (lines 44-48).

In what sense the first library of the automated process is an extension of "traffic" library?

Response

The wording at the beginning of section 3 to clarify that the authors have forked the *traffic* library and implemented functionality on top of it required by this work (lines 107-111).

Assumption have to be mentioned clearly, for instance "always using the maximum available thrust for all departures" only appears in Section 4.3.

Response

The assumptions made are described throughout Section 3.1. The new structure of this section (in response to the following comment) should provide a better overview. Regarding the use of maximum thrust coefficients to estimate thrust for departures, a sentence has been added in Section 3.1 to make this more clear. Note however, that this topic is already extensively discussed in Section 3.3.

The Section 3.1 may be improved given a title to each step (filtering, smoothing...). The part from line 157 to 186 is very detailed compared to the others of the automated process. If it represents a main contribution of the paper, this point has to be mentioned early. If not, the content may be synthesized.

Response

The authors agree with the reviewer regarding the improvements to section 3.1 and have restructured accordingly. Note however, that most of the assumptions made are in the trajectory enhancement part and these should be clearly stated, for which lines 157 to 186 were synthesized only to a certain degree.

Concerning noise and emission results sections, it would be helpful, additionally, to compare the results with those from other studies or models.

Response

Regarding the noise results, comparisons have been added to studies which also used ADS-B or radar data to estimate noise and compared the values with a validation dataset (lines 436-443). For completeness, the extent of the datasets used and the trajectory reconstruction approaches are also mentioned. For emissions results, the author's have not found any literature which uses aircraft trajectory data to more accurately define the time-in-mode spent by each flight in each mode, and also to correct fuel flow and emission indices at each trajectory point based on the conditions observed at the aircraft.

In section "Noise results", in Table 2, how the discrepancy on the results for the station M06 between arrival and departure can be explained?

Response

Text has been added which explains why arrivals noise events at station M06 have the lowest coverage rate (explanation also implicitly clarifies why this is not observed for departures. Lines 378-382).

What are the assumptions mentioned line 360? It is possible to perform a test to rule out one of the 3 causes proposed for discrepancies? The corresponding part describing the differences observed in Figure 4 and 5 has to be reformulated in order to better emphasize the selected cause.

Response

The assumptions made to estimate arrival thrust (for each trajectory point) are detailed in Section 3.1. Text was introduced which revisits these assumptions to ease the reading (line 393). Ruling out any of the mentioned causes would be possible with a validation dataset containing e.g. aircraft net thrust along the flight path (at a resolution at least comparable with the one obtained with ADS-B). While theoretically possible, such datasets are not readily available and usually very limited in scope (e.g. in [3], 13 postal flights of a single aircraft type A319-112 were used). As such datasets are not available to us, we are not able to rule out the mentioned causes. Note that this does not have an impact on the following text, as the discussion around Figures 4 & 5 revolves around noise events produced by departure flights.

At last, how the emission results obtained with the automated process are in agreement with the ones from other studies, potentially on other airports ?

Response

As described in the response above, the authors have not found any literature which uses aircraft trajectory data to more accurately define the time-in-mode spent by each flight in each mode, and also to correct fuel flow and emission indexes at each trajectory point based on the conditions observed at the aircraft.

Citations: Pretto et al. line 23, it would be better to give the reference in this way Pretto et al. in [6], [7] have focused... ,.

Response

We agree with the reviewer and changed the wording.

In citation [14], authors are not mentioned.

Response

Authors added to the citation.

A reference is missing for Doc 29 line 24.

Response

Reference has been included.

The sentence line 42 could be reformulated.

Response

See response to first comment. Introduction section has been rephrased.

Give the details of the acronym FOI line 65 and a reference for this database.

Response

Sentence reformulated to include acronym details.

3.2 Response to reviewer 2

Line 166, a citation is required for taking 90% MLW.

Response

Doc 29 methodology cited.

In Equation 1, γ is the descent angle, not the climb angle.

Response

Wording changed. Specified that the descent angle is negative by convention (according to signs in Equation 1).

Figure 2 should include the runways more clearly or are to be described.

Response

Figure 2 has been improved by adding the runways.

Line 347 establishes threshold for selecting ADS-B noise event by considering absolute LAMAX values of recording stations. These thresholds are selected without comparing and correlating the validation set and ADS-B set. The threshold selected could very well be too high for arrivals, hence the under-prediction. It would be a good idea to include it in this paper, or put it down as something to improve in the future.

Response

The thresholds defined for the ADS-B data are the minimum LAMAX values recorded by the noise stations (i.e. the defined thresholds for the noise stations). Wording in Section 4.1 has been changed for clarity (lines 330-337). In Section 4.3, wording has been changed to explain more clearly the approach and its limitations (lines 369-377). A new suggestion for further work is also made: correlate each ADS-B noise event with a measurement in the validation dataset by using the flight track (including timestamps). The effects of using the threshold methodology to define which noise events are generated with ADS-B data will have a stronger influence in the results especially for noise stations which noise event distribution range includes the threshold value (e.g. M18 departure flights, see Figure 5b). Note however, that lowering the threshold would include more lower LAMAX & SEL values in the ADS-B data distributions, lowering the average values and increasing the under-predictions.

3.3 Response to reviewer 3

While the calculation and processing of the ADS-B data are described in detail, the GRAPE software itself is described only briefly, with limited information provided, unless the reader is familiar with various existing guidelines. A more detailed description is not available elsewhere, such as on the associated GitHub repository, leaving the impression that studying the source code is necessary to understand the underlying methodology. Additionally, the discussion on limitations focuses solely on issues related to the trajectories and thrust settings and neglects the assumptions underlying the calculations performed by GRAPE, such as the LTO cycle. Although these assumptions adhere to international standards, their limitations play a crucial role in assessing the effectiveness and reliability of the noise and emissions estimation method. A more comprehensive discussion of these factors would greatly enhance the estimation of the methodology's overall accuracy and potential impact.

Response

The GRAPE tool abstains from making any assumptions regarding its implemented models. On the contrary, it aims at providing a one-to-one implementation of environmental impact calculation models, and provides the user with as much control as possible regarding both the input data used as well as any control parameters the models may accept. This is presented to the reader in Section 3.2 (lines 221-223). As the major contribution of this work is on the automation of environmental impact calculation (i.e. automating how input data is collected, processed and used by the different models to generate standardized outputs) the multiple models used after this step are not presented. Nonetheless, in section 3.2 it is now more clearly stated that:

- WGS84 coordinate system used for geometric calculations and METAR report closest in time to each operation used when weather data is required.
- fuel flow is modelled according to Doc 9889 by interpolating between 4 LTO points and corrected to altitude conditions with BFFM2.
- noise is modelled with Doc 29 and the SAE-ARP-5534 atmospheric absorption model is used.
- FOA 4 method used to estimate nvPM EIs in case they are missing from the EEDB.
- for the first emissions run, segments obtained from ADS-B data are used and the BFFM2 is used to obtain gas pollutant EIs based on fuel flow and corrected to weather conditions observed at the aircraft.
- for the second emissions run the LTO cycle is used, disregarding any trajectory data or flight specific conditions.

Further detailing the assumptions on which these models are constructed (especially the Doc 29 for noise modelling and the BFFM2 for emissions) and which are not controllable by the user is in the authors opinion not required as these are well established standards.

Lack of comparison to existing practice(s) of noise modelling: While the methodology for computing noise data and its comparison to actual measurements is reasonably well-detailed, it is unclear whether the authors' approach improves upon or is at least comparable to established methods, such as those described in ICAO DOC 9911 or comparable studies the authors have considered in the introduction (sources 3 and 4). Including a comparison with existing methodologies would add valuable context for evaluating the effectiveness and relevance of the approach developed by the authors. This would help clarify its role within the state of the art for aircraft noise estimation.

Response

The authors fully agree with the reviewer. Comparisons have been added to studies which also used ADS-B or radar data to estimate noise and compared the values with a validation dataset (lines 436-443). For completeness, the extent of the datasets used and the trajectory reconstruction approaches are also mentioned. Note that source 4 does not provide results regarding the comparison with measurement values. In section 5, text has been added which details the addition of other trajectory reconstruction methods to further work (lines 529-533). Note that the Doc 9911 is in content identical to Doc 29. The noise estimation component of the automated process proposed is already using the Doc 29 noise model. In section 5, we propose analysing how using the other components of Doc 9911/Doc 29, i.e. the performance and trajectory modelling, would impact the results obtained (both for noise and emissions).

I believe the authors' approach has significant potential to improve emissions estimations based on the LTO cycle. Incorporating real trajectories and weather conditions, especially for thrust computation, presents a notable step forward. However, some significant shortcomings remain with the

chosen approach. For example, while weather conditions are partially accounted for in thrust calculations, their broader impact on emissions and their impact is not addressed. The primary focus of the discussion is on comparing calculated values for time spent in arrival (LTO: approach) and departure (LTO: takeoff and climb) modes. An additional difference lies in the thrust settings, which are approximated in the authors' method but fixed in the LTO cycle. While these differences highlight deviations in real trajectories worth observing, the resulting variations in emissions are a logical consequence of these adjustments and, as such, not inherently significant. Essentially, the authors state that "we change variables that impact emissions, and we observe changes in emissions," without providing any context as to whether these results are meaningful or valuable. A comparison with in-situ measurements, as was done with noise, other established approaches to emission modelling (e.g., the methods outlined in ICAO DOC 9889), or even a discussion of the broader implications of these findings, would add much-needed context and elevate the discussion of emissions in the paper.

Response

The authors agree with the reviewer and have completely restructured the discussion of the emissions results. The impact of each of the emissions calculation stages using ADS-B data compared to the LTO cycle, namely differences in time below 3000 ft, fuel flow and emission indices are now explicitly presented and discussed. The significance of each stage on the estimated emissions compared to the LTO cycle is discussed based on the total impact on pollutants estimated across the whole year. This discussion is equivalent to comparing the ICAO Doc 9889 simple approach and multiple stages of the sophisticated approach. The comparison of in-situ measurements to modelled values for emissions is only possible through the modelling of emissions dispersion and estimation of concentration values at different locations, which is (as now more clearly stated) outside the scope of this work.

It is not entirely clear what kind of weather data has been used, in what way. The text discusses "airport and time specific weather data" (line 59), "Airport weather observations" (Table 1), or "the weather report closest to each trajectory point" (line 142) while the schematic in figure 1 contains METAR. I assume this means that METAR reports were used throughout. Is this also used to calculate true air speed for the trajectory segments near 3000ft altitude?

Response

Wording has been changed to make it clear to the reader that METAR reports are being used throughout the paper. As described in Section 3.1, METAR reports are being used to obtain weather information when required for feature enhancement, i.e. also to determine true airspeed.

Setting the taxi time to zero really provides data only "near" airports and not "at" airports. While the authors never claim the latter and discuss arrival and departure throughout the paper, I believe it would be helpful to explicitly mention that traffic at the airport itself is not considered in this estimation.

Response

The reworked introduction states more clearly the objectives and scope of this work, including that it focus is on estimating environmental impacts *in the immediate vicinity of airports*.

It is not explained and how the threshold levels for noise described in line 312 and following are chosen. Where do these values come from and why is there no difference between day and night for two stations?

Response

Text has been added to section 4.1 to explain how the thresholds are obtained (lines 331-337). In section 4.3, text has been added which explains how these thresholds are used to simulate what constitutes a noise event in the estimated data (lines 369-377). Furthermore, an improvement on this approach is suggested for further work.

The flight results show that approximately 81.5% of the number of flights were identified by the developed method. It is not clear to me if and how these roughly 20% difference are considered when comparing the modelled noise events with the measured ones.

Response

This work goes a step further, as not only differences between flights recorded by the ADS-B network and the airport are identified, but also the differences between number of estimated noise events and measured noise events (see Table 2). These differences become further evident in the histograms comparison (Figures 3-5). In the discussion of the year L_{eq} results (Table 3), these differences are also explicitly stated (lines 425-427).

4. Review - round 2

4.1 Reviewer 1

Thank you for the detailed responses to my previous comments. I have some additional revisions to suggest.

1. On page 10 and line 367, the location of station M11 is deemed problematic. But with the addition of runways in Figure 2, I see that M11 is a good location (lateral to the runway). Since it should be able to capture the noise events, I don't see how the location might be the reason for errors. Can you be more specific?

2. On page no. 11 and line 411, the use of similar shapes doesn't seem entirely accurate. The subfigures in Figure 4 show a Gaussian curve in the validation set, but the ADS-B set is primarily bimodal. Can you please elaborate on your reasoning?

3. On page no.12 and line 419, it's eluded that there is a problem with the quality of NPD tables. If the quality of SEL NPD tables in the ANP database is incorrect, then why would LAMAX NPD tables be correct? Could you please elaborate? Maybe refer to works on calibrating NPD tables by van der Grift et al. (Aircraft Noise Model Improvement by Calibration of Noise-Power-Distance Values Using Acoustic Measurements | Aeroacoustics Conferences)?

4. Page 13 and line 423 describe the year noise equivalent sound levels L_{eq} . I did not fully understand this metric. Could you please include the mathematical equation or a reference for this?

5. Page 13 and line 431 explain the reasoning behind the improved accuracy for departure operations. However, the influence of this canceling effect is not sufficiently explained. For example, more information about the other potential problems due to the counteraction would complete this assessment. Assuming maximum thrust for departures in noise modeling is not common anymore.

a) Schwab et al. (<https://arc.aiaa.org/doi/10.2514/1.C035779>) have described the trends in N1% of departure procedures using FDR data.

b) N1% is also calculated acoustically, as done by Merino-Martinez et al. (<https://arc.aiaa.org/doi/10.2514/1.C034849>).

c) More recently, Meister has used machine learning techniques to estimate the power settings (<https://arc.aiaa.org/doi/epdf/10.2514/1.C037619>).

Integrating these methods with the automation process in this paper is an interesting topic for future work.

4.2 Reviewer 2

The changes introduced by the authors add significant context and have overall improved the quality of the paper. I would particularly like to highlight the improvements made to the introduction, which now clearly states the aim and scope of the paper. This revision has greatly enhanced the clarity and focus of the study.

That said, I still find the discussion of the limitations regarding the calculation of emissions to be insufficient. While some aspects are mentioned in Section 3.3 – such as the uncertainties related to take-off weight – other important points remain unaddressed. For example, although the use of METAR data is certainly an improvement over relying solely on ISA conditions, it also has its own limitations that should be acknowledged. Likewise, the limitations of the databases used as inputs for GRAPE should be addressed in more detail.

The newly added discussion of emissions is a valuable addition that significantly strengthens the paper, and I found this part particularly interesting to read. However, I believe it would benefit from a more explicit reflection on the method's limitations and its implications. Without comparison to “real-life” data, whether measured or modeled, it remains unclear whether the chosen approach offers a meaningful improvement over standard LTO-cycle-based estimations. While I understand that incorporating such a comparison may be beyond the scope of the current study – or even unfeasible for this particular case – a critical discussion of these limitations would still be a welcome addition to better assess the method's applicability and potential impact.

I recommend the paper for publication, provided the authors include a critical discussion of the limitations mentioned above

5. Response - round 2

We would like to thank the reviewers for their comments and insights. The paper has been updated and the response to each comment is provided below.

5.1 Response to reviewer 1

On page 10 and line 367, the location of station M11 is deemed problematic. But with the addition of runways in Figure 2, I see that M11 is a good location (lateral to the runway). Since it should be able to capture the noise events, I don't see how the location might be the reason for errors. Can you be more specific?

Response

It is stated in Section 4.3 that the discrepancies observed for the discarded stations are of significant magnitude and most likely due to an incorrect representation of the stations location (lines 374-376).

On page no. 11 and line 411, the use of similar shapes doesn't seem entirely accurate. The subfigures in Figure 4 show a Gaussian curve in the validation set, but the ADS-B set is primarily bimodal. Can you please elaborate on your reasoning?

Response

The authors agree with the reviewer regarding the differing distributions and have incorporated additional text to describe this and provide a potential explanation for the underlying cause. (lines 411-419).

On page no.12 and line 419, it's eluded that there is a problem with the quality of NPD tables. If the quality of SEL NPD tables in the ANP database is incorrect, then why would LAMAX NPD tables be correct? Could you please elaborate? Maybe refer to works on calibrating NPD tables by van der Grift et al. (Aircraft Noise Model Improvement by Calibration of Noise-Power-Distance Values Using Acoustic Measurements | Aeroacoustics Conferences)?

Response

The values in the NPD tables are empirical and specific to each aircraft type. Variations between individual aircraft operating within the fleet mix at Cologne Bonn Airport and their corresponding NPD aircraft type may be more pronounced for one of the two metrics, leading to differing results. In fact, assuming all other factors remain constant, such as enhanced trajectory, atmospheric parameters, and noise calculation methodology, differences in the noise source description are the most plausible explanation. A reference has been added to a paper that observed a similar discrepancy for departure flights (as opposed to arrival flights) when comparing noise estimates derived from the Doc29 methodology with measured values (pages 437-438). The paper cited by the reviewer exhibits a similar trend for Oslo in Table 3; however, it is not explicitly referenced, as this is not a primary conclusion or focus of the study. Additionally, the NPD table calibration methodology outlined in the referenced paper falls outside the scope of this work, as it relies on noise measurement data specific to each airport. Such data is often limited or unavailable, making it impractical to integrate into the automated process.

Page 13 and line 423 describe the year noise equivalent sound levels L_{eq} . I did not fully understand this metric. Could you please include the mathematical equation or a reference for this?

Response

The name has been changed to clearly indicate that it refers to the (common) equivalent continuous sound level $L_{eq} = 10 \log_{10} \left(\frac{1}{T} \int_0^T 10^{\frac{L(t)}{10}} dt \right)$ with $T = 1$ year.

Page 13 and line 431 explain the reasoning behind the improved accuracy for departure operations. However, the influence of this canceling effect is not sufficiently explained. For example, more information about the other potential problems due to the counteraction would complete this assessment. Assuming maximum thrust for departures in noise modeling is not common anymore.

a) Schwab et al. (<https://arc.aiaa.org/doi/10.2514/1.C035779>) have described the trends in N1% of departure procedures using FDR data.

b) N1% is also calculated acoustically, as done by Merino-Martinez et al. (<https://arc.aiaa.org/doi/10.2514/1.C034849>).

c) More recently, Meister has used machine learning techniques to estimate the power settings (<https://arc.aiaa.org/doi/epdf/10.2514/1.C037619>).

Integrating these methods with the automation process in this paper is an interesting topic for future work.

Response

The limitations of the developed automated process regarding thrust estimation, and potential improvements, are extensively discussed in Section 3.3. The potential use of machine learning models is also discussed (including a reference to previous work by Meister). References a) and b) use approaches which require a per-airport calibration and/or not widely available data, for which they are outside the scope of this work. The text at the end of Section 4.3 lists the discrepancies in the data used to calculate Leq between ADS-B data and the validation dataset, i.e. differences in the number of noise events, as well as the potential influence of other factors from the data and methodology used. However, a more detailed analysis to quantify the impact of specific parameters and assumptions on the results would require a more in-depth study, which falls beyond the scope of this work.

5.2 Response to reviewer 2

That said, I still find the discussion of the limitations regarding the calculation of emissions to be insufficient. While some aspects are mentioned in Section 3.3 – such as the uncertainties related to take-off weight – other important points remain unaddressed. For example, although the use of METAR data is certainly an improvement over relying solely on ISA conditions, it also has its own limitations that should be acknowledged. Likewise, the limitations of the databases used as inputs for GRAPE should be addressed in more detail.

Response

Text has been added to Section 3.3 which addresses further limitations of the automated process developed, namely the use of METAR data and the detail level of the aircraft abstraction (lines 262-266 and 288-294)

The newly added discussion of emissions is a valuable addition that significantly strengthens the paper, and I found this part particularly interesting to read. However, I believe it would benefit from a more explicit reflection on the method's limitations and its implications. Without comparison to “real-life” data, whether measured or modeled, it remains unclear whether the chosen approach offers a meaningful improvement over standard LTO-cycle-based estimations. While I understand that incorporating such a comparison may be beyond the scope of the current study – or even unfeasible for this particular case – a critical discussion of these limitations would still be a welcome addition to better assess the method's applicability and potential impact.

Response

We agree with the reviewers statement and added text at the end of section 4.4 (lines 517-524) which more clearly outlines the limitations in the interpretation of the results obtained and suggests potential steps to overcome them.