JOAS

EDITORIAL

Reviews and responses for

Investigation of Point Merge Utilization Worldwide Using Opensky Network Data

Authors: Henrik Hardell, Tatiana Polishchuk, and Lucie Smetanová

Reviewers: Max Li, Ryota Mori, Enrico Spinielli

Editor: Xavier Olive

1. Original paper

DOI for the original paper: https://doi.org/10.59490/joas.2023.7218

2. Review - round 1

2.1   Reviewer 1

The authors utilize OpenSky aircraft trajectory data in order to test out a method for detecting whether or not a flight used PM procedures, and then reporting out the utilization ratio for an airport where there are active PM procedures. Although it is a bit surprising to me that such percentages are not standard in terms of reported metrics (e.g., if we had access to filed flight plans, would we be able to easily calculate this?), I believe this study is a good example of simple arrival procedure KPIs that can be calculated using open-source trajectory data. I have some comments/suggestions for the authors below.

1. Although it may be clear to people who are familiar with PM procedures at the airports listed in the paper, it is not immediately clear whether or not the authors examine PM procedures for different airport flow configurations. For example, for DUB, the depicted PM in Figure 1 appears to be for a west flow configuration. I assume (although I may be wrong) that there is an analogous PM procedure for east flow operations at DUB. I don't think the authors need to include the STAR procedure diagrams for every PM procedure at the airports that they analyze in this paper, but I would suggest mentioning explicitly that, e.g., Figure 1 depicts only a subset of PM procedures at each airport.

2. Would there be any way to construct a set of validation flights to check the PM utilization KPI? For example, would it be possible for the authors to, e.g., collect a set of filed flight plans for flights that were last filed to utilize a PM approach, and then compare how their estimation methodology for whether or not a flight used a PM approach performs on this set of validation flights? I am not sure if such a validation set could be constructed, but it seems to me that that might be a better way to check the effectiveness rather than visually inspecting the trajectories that were classified as PM or non-PM (e.g., the "correctness check" in Figures 3 and 4)

3. I would suggest that the authors format Figure 5 to be more similar to a standard empirical CDF – starting at 0% of the PM sequencing leg, and then increasing to 100%, and seeing how much this

captures in terms of the percentage of flights that did fly a PM procedure.

4. The detection methodology, maybe unsurprisingly so, appears to be pretty sensitive to the size of the circular catchment area. Related to my previous comment regarding potentially constructing a validation set, such as set would also be helpful in determining an "optimal" catchment size for each airport, since there may be airport-specific differences in how to most successfully capture PM procedure flights.

2.2   Reviewer 2

This paper focuses on PM procedure, and the arrival flows at various airports are analyzed. Although the authors successfully identified the PM flights, I feel the contribution of this paper is insufficient for the journal publication in the current form. I suggest that the authors should do additional work and provide in-depth analysis using the obtained data. The following is the list of the suggested additional work. Please consider the inclusion of these topics for the paper to be published.

1. The authors focus on PM utilization based on the waypoints. However, the distance from the initial PM waypoint will also be a factor. The length of the segment between waypoints differs among airports, and additional insights might be obtained.

2. In the PM utilization, some aircraft seem to fly over the full arc, but are there any aircraft flying beyond the full arc? Something interesting might be obtained.

3. In Tables 3-9, if I understand correctly, each line indicates each PM system, not the PM sequencing leg. (A single PM system includes multiple PM sequencing legs.) If so, it may be interesting to see the difference in leg usage among the legs. (some legs may be prioritized, for example) Also, traffic volume (or traffic volume at peak hours) may also affect the result.

Also, there are some minor comments.

1. The aspect ratio of the most pictures is wrong. For example, in Fig. 3, if the aspect ratio is correct, the red circles must be a circle, but some are ellipses. Especially, Fig. 3 (c) and (e) look strange.

2. In Tables 3-7, "All PM" is not needed, because not all Tables have "All PM" rows.

2.3   Reviewer 3

Note: Given no pre-prepared PDF was provided, the line numbers refer to the PDF produced by dropping the zip in Overleaf which I attach to this review.

Interesting paper. It seems to be the first of a series where deeper analysis will follow, the PM length utilization metric is not meaningful per se and could be useful with other indicators to assess the efficacy/efficiency of an airport.

Lines: 45-47, Dublin is missing the number of movements. Lines 43-59 could possibly be replaced by a table with cols like airport (continent?), movements, (PM) since, (PM) description.

Figure 3: not clear what is caught by the catchment algorithm and what is not. Lines 115-143 mention point IDs (SIVNA, KOGAX, LUTIV, BR635...) which are not in Figure 3, and it is difficult to follow the text

Section 2.4.1/lines 142-146: the correctness check is done visually, but I guess it was then quantified how many PM flights were not identified as such and as well as how many non-PM flights were caught. These figures could help in assessing the correctness of the approach.

Lines 166-174: the text seems to hint at the percentage utilization of the length of the PM system as a measure of the proper functioning of PM. It should be stressed that full utilization of the PM length

shows that there are congestion and sequencing problems. If full-length utilization is sustained for long stretches then the airport has difficulty managing arrivals. So length utilization should be accompanied with more contextual information to assess the quality of PM utilisation.

For example, what is the assessment for Bergen, lines 173-174? Similarly, lines 198-202 should stress that length utilization is just one metric to be combined with other contextual information/metrics.

Reproducibility: I cloned the repo and read/tried some of the code. I created an environment as follows (somewhat following README):

```
mamba create -n lucie_osn23 -c conda-forge python=3.10 matplotlib numpy pandas pyproj shapely
```

I gave a try to example_code_to_determine_runways.py but got immediately stuck: it is not at all clear where to get the data from and the README doesn't provide enough details.

It says: "Before running the code, you need to specify your path to the input/output data."

From the code

```
DATA_DIR = os.path.join('\Data', airport_icao)
DATA_DIR = os.path.join(DATA_DIR, year)
```

which when run gives 'Data/EIDW/2022' so I guess (but I could be mistaken) I should download the historical state vector from OSN for the relevant period.

I feel like a script to prepare (a subset of) the data (why not a Zenodo dataset specific for the paper?) is essential to allow the execution end-to-end of the analysis in the paper.

I would also suggest avoiding OS-specific paths/settings.

It would be helpful to have a description of the structure of the input and output data directories and relevant files. Finally, I would suggest having the examples working for a subset (i.e. one) of the airports (and maybe for a reduced time period).

3. Response - round 1

3.1 Response to Reviewer 1

Although it may be clear to people who are familiar with PM procedures at the airports listed in the paper, it is not immediately clear whether or not the authors examine PM procedures for different airport flow configurations. For example, for DUB, the depicted PM in Figure 1 appears to be for a west flow configuration. I assume (although I may be wrong) that there is an analogous PM procedure for east flow operations at DUB. I don't think the authors need to include the STAR procedure diagrams for every PM procedure at the airports that they analyze in this paper, but I would suggest mentioning explicitly that, e.g., Figure 1 depicts only a subset of PM procedures at each airport.

 > **We agree with this suggestion, and we added an explanation sentence in the Airports section where we describe the Figure.**
> **Changes in manuscript: Added sentence in lines 42-44.**

Would there be any way to construct a set of validation flights to check the PM utilization KPI? For example, would it be possible for the authors to, e.g., collect a set of filed flight plans for flights that were last filed to utilize a PM approach, and then compare how their estimation methodology for whether or not a flight used a PM approach performs on this set of validation flights? I am not sure if such a validation set could be constructed, but it seems to me that that might be a better way to check the effectiveness rather than visually inspecting the trajectories that were classified as PM or non-PM (e.g., the "correctness check" in Figures 3 and 4)

 > **This is a very interesting suggestion, however, there are unfortunately no suitable flight plan data available for the flights in our study. Moreover, we believe that the Point Merge sequencing legs usage is not present in the original flight planning, as they are used as an instrument for the air traffic controllers for the tactical deconfliction. The default aircraft route in the absence of congestion is planned as a direct route from the start of the sequencing leg arc to the merge point (the shortest route). Aircraft are instructed to enter the sequencing leg arcs when they need to wait for the other aircraft to land. Therefore, the Point Merge arc's occupancy is an indication of the congestion during the given moment. That is why we think it is important to construct a metric for quantification of the PM occupancy based on the actual (historical) aircraft trajectories.**

I would suggest that the authors format Figure 5 to be more similar to a standard empirical CDF – starting at 0% of the PM sequencing leg, and then increasing to 100%, and seeing how much this captures in terms of the percentage of flights that did fly a PM procedure.

 > **We appreciate this comment and confirm that the changed Figure according to the CDF standards fits better to the paper. Changes in manuscript: Changed Figure 5 in PM Utilization Experimental Results section.**

The detection methodology, maybe unsurprisingly so, appears to be pretty sensitive to the size of the circular catchment area. Related to my previous comment regarding potentially constructing a validation set, such as set would also be helpful in determining an "optimal" catchment size for each airport, since there may be airport-specific differences in how to most successfully capture PM procedure flights.

 > **As we stressed in response to comment number 2, there are unfortunately no flight plans available at the moment. We suspect that even if the flight plans were available, judging from our experience with DDR data, wouldn't be detailed enough and wouldn't contain the planning so close to the final approach. We believe the 'optimality' must be detected experimentally and the size of the catchment area may be different for different airports. We plan to include more extensive sensitivity analysis in the future journal article submission.**

3.2   Response to reviewer 2

The authors focus on PM utilization based on the waypoints. However, the distance from the initial PM waypoint will also be a factor. The length of the segment between waypoints differs among airports, and additional insights might be obtained.

 > **We are thankful for the comment and are considering analysing the effect of the different PM geometries on the PM efficiency metrics in the future. However, in this work, the presented utilization KPI is not intended to capture the actual distance flown along the arc, but rather the proportion of the arc utilized.**
> **Changes in manuscript: We added an explanation sentence to the updated version of our paper in subsection PM Utilization of section KPIs on lines 96-99.**

In the PM utilization, some aircraft seem to fly over the full arc, but are there any aircraft flying beyond the full arc? Something interesting might be obtained.

 > **Interesting question. During our analysis we have never observed such a phenomenon, moreover, we believe that such action is restricted by the Point Merge procedure design itself.**

In Tables 3-9, if I understand correctly, each line indicates each PM system, not the PM sequencing

leg. (A single PM system includes multiple PM sequencing legs.) If so, it may be interesting to see the difference in leg usage among the legs. (some legs may be prioritized, for example) Also, traffic volume (or traffic volume at peak hours) may also affect the result.

> **This comment brings up a good point and helpful suggestion to include in our work. The tables regarding PM Utilization indicates the number for each sequencing leg separately in most cases. The only exceptions are Oslo Gardermoen airport and the Eastern PM system at Dublin airport, which operate conventional PM design and thus the PM flights were easy to identify for the whole PM system at once. Most of the other airports have some sequencing-leg-specific features which need to be addressed separately in the catchment algorithm. To address the comment, we included Table 5 covering the PM Usage values for each of the sequencing legs separately.**
> **Changes in manuscript: We added Table 5 in the section Experimental Results – PM Usage and explanation sentence on lines 174- 177.**

The aspect ratio of the most pictures is wrong. For example, in Fig. 3, if the aspect ratio is correct, the red circles must be a circle, but some are ellipses. Especially, Fig. 3 (c) and (e) look strange.

> **This is a valid point. In the original version, we aimed to show the whole PM system parts and most of the important parts of the flight trajectories which caused the squawked circle catchment areas in the pictures. We fixed this error in the updated version of the paper.**
> **Changes in manuscript: Updated some subfigures in Figure 3.**

For tables 3-7, "All PM" is not needed, because not all Tables have "All PM" rows.

> **Thank you for the valuable comment, we would like to explain our reasoning here. We included the 'all PM' rows to all tables of PM Utilization results for airports which have two or more comparable (the same number of segments) size PM system arcs. In the 'all PM' row, the PM Utilization values are calculated based on the accumulated values from each of the contributing sequencing legs. First, we sum up the trajectories passing each segment regardless of the sequencing leg location and then we calculate the percentage from that. We think this additional information gives an overview of how the airports work with the overall PM systems.**
> **Changes in manuscript: We added an explanation to subsection PM Utilization of Experimental Results section on lines 184-188.**

3.3   Response to Reviewer 3

Lines: 45-47, Dublin is missing the amount of movements. Author's response:

> **That is a very valid comment, we included the number in the updated version of the paper.**
> **Changes in manuscript: Added value in section Airports, lines 49 and 50.**

Lines 43-59 could possibly be replaced by a table with cols like airport (continent?), movements, (PM) since, (PM) description.

> **We appreciate this comment and add such a table to the updated version of our paper. We agree that the table improves the readability of the section.**
> **Changes in manuscript: Added Table 1 in line 71.**

Figure 3: not clear what is caught by the catchment algorithm and what is not.

> **This is a valid point, we described it better in the updated version. Figure 3 shows only**

the trajectories attributed to the Point Merge system: the ones caught by the catchment algorithm. We analyze the trajectories which were not identified as Point Merge flights separately to understand the false-negative and false-positive mistakes.
> Changes in manuscript: We attached a sentence explaining the results in Figure 3 and on lines 113-115.

Lines 115-143 mention point IDs (SIVNA, KOGAX, LUTIV, BR635...) which are not in Figure 3 and so it is difficult to follow the text

> Very good comment. We altered the text to increase the readability. We marked the circle catchment areas with colors in Figure 3 and explained the colors in the text for each of the pictures in the Figure. The circles colored in red are the main catchment areas at the start of each sequencing leg. The blue colored circles are additional catchment areas to accommodate the incoming traffic flow, or to correct the catchment algorithm in cases of consistent malfunction due to the design of the PM system. The green colored circle is added to the Bergen airport PM system to accommodate the traffic incoming to the sequencing leg but to filter out flights not performing the PM procedure.
> Changes in manuscript: We changed the pictures in Figure 3 and attached an explanation to each in the section Catchment algorithm on lines 116-156.

Section 2.4.1/lines 142-146: the correctness check is done visually, but I guess it was then quantified how many PM flights were not identified as such as well as how many non-PM flights were caught. These figures could help in assessing the correctness of the approach.

> Yes, that is true. We had those numbers calculated and partially used them in the PM Usage calculation however, we haven't described them explicitly in the paper. In connection with your comment, we attached Table 3, which provides the information about the number of all arriving flights in the dataset, the number of identified flights as the PM flights and the number of false-positive flights for each airport. In future work, we plan for a more extensive sensitivity analysis in order to identify the radius of the catchment area circle which minimizes the number of false positive flights, but at the same time, still catches the actual PM ones.
> Changes in manuscript: Attached is Table 3 and an explanation sentence on lines 161-166 and 177.

Lines 166-174: the text seems to hint at the percentage utilization of the length of the PM system as a measure of the proper functioning of PM. It should be stressed that full utilization of the PM length shows that there are congestion and sequencing problems. If full-length utilization is sustained for long stretches then the airport has difficulty managing arrivals. So length utilization should be accompanied with more contextual information to assess the quality of PM utilisation. For example, what is the assessment for Bergen, lines 173-174? Similarly, lines 198-202 should stress that length utilization is just one metric to be combined with other contextual information/metrics.

> We appreciate this comment, and we acknowledge it in the text.
> Changes in manuscript: An appended sentence in the section Conclusions and Future Work on lines 228-231.

Difficulties in reproducibility

> Thank you for the valuable comment. As the JOAS Journal and Conference Proceedings are our first experience with open access to the data repositories and codes, we did our best to provide the necessary parts. However, we agree our GitHub repository needs to be upgraded and we are working on that. We are also adding a smaller demonstrational subset

which can be run easily.
> Changes in manuscript: Provided major changes in the GitHub repository.

4. Review - round 2

4.1 Reviewer 1

I thank the authors for their time and efforts in addressing the comments. I am happy to recommend acceptance.

4.2 Reviewer 2

The authors adequately respond to the questions, and I feel that the paper is ready for publication.

4.3 Reviewer 3

The comments I submitted have been nicely implemented: thank you to the author(s). In order to reach paper reproducibility, there are some improvements in the Github repo but I think these can be handled offline: I submitted a few Github Issues to track them. The author(s) demonstrated a professional commitment to reproducibility and I am sure she (they) will submit the relevant fixes. These are no showstoppers to go ahead with the paper publication.

Congratulations on the nice paper.