

EDITORIAL

Reviews and Responses for Investigation of Point Merge Utilization Worldwide Using Opensky Network Data

Authors: Henrik Hardell, Tatiana Polishchuk, and Lucie Smetanová

Reviewers: Raúl Sáez, Enrico Spinielli, and Aymeric Trzmiel

Editor: Junzi Sun

1. Original paper

DOI for the original paper: <https://doi.org/10.59490/joas.2025.7729>

2. Review - round 1

2.1 Reviewer 1

Overall, I think the article is pretty much ready for publication, but the (sample) code should be fixed. On the paper, I have the following comments:

Line 120: the sentence with 'granulated' should be changed to something like "...trajectory data in the form of state vectors at 1s resolution."

Figure 3: Add a legend for the meaning of shapes (4) and colors (4) [aren't they representing the same concept? If yes, maybe shapes are good enough].

Figure 5: Annotate the plots with the waypoints mentioned in the relevant text, i.e., NONVA, GODLU, BABKU...

Lines 210-211: It is not clear how four waypoints (LI739, LI760, LI725, LI748) can be marked by two red circles...again, annotating the plot with the waypoints could help.

Line 270: "We horizontal and" is missing a verb.

Table 3: Add a column with

Table x: Add a table similar to Table 3 but for False Negatives.

Table 6: It is not clear what Airport Width, Airport Length, or Rectangle means. My understanding, from lines 299-300, "smallest rectangle around the TMA," is that an axis-aligned geographical bounding box is taken for each airport, then Airport Width (in km) = $\max(\text{lon}) - \min(\text{lon})$, rectangle (km^2) = area of the bonding box, Airport Length (km) = $\max(\text{lat}) - \min(\text{lat})$. Please clarify. (BTW, the smallest rectangle could well be one that is not axis-aligned but rotated.)

Lines 289-301: I find this check on correlation with quantities such as airport length, airport width, and airport area quite misfit.

Lines 313-314: Too many "only" instances.

Lines 328-333: The number in Table 7 does not seem to indicate that the two algorithms catch approximately the same number of flights; the differences are too big! (apart from EIDW East).

Code: Output directory in 'Small example' / could be added in the repo (you can just add a "hidden" empty file to have git keep it).

Code: I advise against directory or file names containing blanks.

Code: Spellcheck README.md

Code: The 06_VP_ENGM.py file fails to run: wrong ident error.

Code: The subsequent example scripts are stuck.

Code: Script 10 has a syntax error.

Please fix the code and make sure at least the sample scripts work out of the box.

2.2 Reviewer 2

In this paper, the authors use a "catchment" algorithm to detect flights flying point merge procedures. This algorithm is based on two steps, focusing on the lateral and vertical profile of each flight. The authors also present a sensitivity analysis of some of the parameters used in this algorithm.

The paper is well-written and easy to follow, with clear explanations of the proposed methodology.

Still, I have some comments:

1) Regarding the correctness checks, I would like the authors to elaborate further on what they mean by "visual observation". Did the authors just plot all the trajectories and then look for false positives and false negatives manually? How did the authors identify the exact flights that were not identified well by their catchment algorithm?

2) Regarding Table 7, the authors claim in Line 328 that there are just big differences between Oslo Airport and St. Petersburg Airport. However, in Table 7, EIDW (West), ENBR, ENGM, RKSI, and ULLI all have big differences in the number of flights detected between both algorithms. The authors should also acknowledge this fact.

Any idea regarding why these big differences are happening?

Furthermore, are the flights detected by each algorithm the same flights? The quantity might be similar, but the specific flights detected might be different.

3) What are the implications of having the PM usage and utilization rates the authors present in Section 5? Were the PM systems in these airports badly designed? Are the ATCs not willing to use these new procedures? Are there any other reasons? I would like the authors to elaborate further on this topic.

4) Miscellaneous/typos:

4.1) Line 142: "unsafe"—I am not sure what the authors mean by this. . .

4.2) Line 149: "to the point, when the aircraft" — comma not needed

4.3) Line 159: "and if yes, the algorithm proceeds to the second step, we analyse the vertical profiles." — probably needs a rephrasing: "and, if so, the algorithm proceeds to the second step, in which we analyse the vertical profiles."

4.4) Figure 3: Font size should be increased to improve readability.

4.5) Line 266: The "False negatives" title should also be in bold or italic, like the false positives title.

4.6) Line 270: "We horizontal and.." Please correct this.

2.3 Reviewer 3

Overall:

The paper and study are well documented, with numerous figures and reports on a lot of work for data cleaning and for the method to quantify the defined indicators. However, the overall aim and objective of the study are not very clear to me. What is the purpose of measuring the use of the sequencing legs? Which problem does it solve? How does it contribute to aviation science? What are the operational impacts of the results?

No link is made between the utilization of PM and the traffic demand, delay absorption capacity, or functioning regime (e.g., adapt point merge legs to the traffic demand). Indeed, the number of aircraft does not define the traffic demand (as such, the capacity needed). A notion of time/rate should be added at the entry and exit of the PM systems.

No link is shown between the various configurations of PM (overlapping leg, one/dual runway) and the results of the utilization.

What do the enhancements perform, compared to the previous research presented at the 11th Open-Sky symposium, bring in terms of performance? E.g., accuracy, precision, recall F1-score. The same applies for comparison with the algorithm from the traffic library.

Page 1

Abstract: "...the geometry...PM or merging points impact the associated trade-offs." Could you please clarify what the trade-offs you mentioned are?

Page 4

In addition to the airport info provided, interesting info to add would be how the point merge is used according to runway use: e.g., In the case of a 1 PM system for 2 runways, is it used in an independent/dependent parallel, staggered approach? The same applies if two systems for 2 runways...In addition to traffic load and demand, the final rate delivered at the airport is also important for calculating the capacity.

Page 6

Table 2: I understood that the number of flights corresponds to the results once data is cleaned/-filtered (removing go around, holdings...). Could you please clarify what the percentage of flights removed from the initial dataset is?

Line 142: There might be a wording error in that sentence: "This metric indicates the frequency of PM procedure unsafe during the period under consideration." What does "unsafe" mean here?

Chapter 4 on Catchment algorithm:

A lot of tuning has been made to capture the relevant flight, including sensitivity analysis for the radius sizes that have to be adapted to each airport. In addition, manual observation is still required to distinguish false positives. If the purpose is to measure the distance flown along the sequencing legs, why not use a direct reference trajectory from an entry to an exit point/or delimited area per flow and per airport/runway and measure the actual distance flown between all tracks and these reference trajectories? The difference should correspond to the additional distance flown along the

leg, shouldn't it? As (most of) the flights are direct to the merge point when leaving the legs. Then, as the length of the legs is known, derive the percentage of leg use.

To me, using the sequencing leg points is pointless, as, no matter the number of points defining the leg, the measure indicating the maximum delay absorption capacity is the total length of the legs.

All the figures with tracks and procedures are distorted. Could you please re-do them with the correct projections or orthonormal frame? Here, the equidistances between the sequencing legs and the merge point, which is the basic principle of the PM procedures, are not visible in the figures.

Chapter 4.0.4

Column titles of table 6 are misleading. PM size is (if I understood well) the length of the sequencing legs. What are Airport width/length and Rectangle? (is it TMA surface?) Some are in NM; others are in km. I would suggest removing it as it does not bring added value, as no correlation is found.

Chapter 4.1

Table 7: To enable comparison between the catchment algorithm and the Traffic library, I would appreciate dedicated indicators (accuracy, recall, precision,...) instead of raw difference numbers. At the end, what is the most accurate/efficient algorithm?

Line 324-333:

To me, Bergen is not the only airport with fully separated PM. EIDW for runway 10L/R (PM East) could be considered as such (as defined in Table 1).

How can you explain such a wide variation of performance ranging from -98% to +118%? So, from catching nothing to catching double! With such values, it can't be said that the algorithms "catch approximately the same number of flights," except !! And once again, it is not because they catch the same number of flights that they perform equally; additional indicators such as accuracy and recall are needed. This entire section should be reworked.

Figure 9: Differences or similarities between (a) and (b) and between (c) and (d) are not visible (very small, not the same scale...). Displaying tracks not recognized by one algorithm but caught by the other (or the opposite) would be more helpful to understand.

Chapter 5.2

Line 358: Could you please further explain what "PM system not unified" means?

Tables 10, 11...:

All these tables are quite cumbersome to read and to compare, even more so with sub-tables according to radius. Why not use only one radius value (revised or initial), e.g., the one that performs better? A Pie chart will ease the visualisation of the radius.

Please make them consistent: ranges/bins of percentage are different for each table: 20%, 50%, 25%, 1/3....

At the end, what I'm interested in is, what the PM utilization in terms of distribution: how many go direct (early direct or direct "just at leg entry"), what is the median, quantile, percentile....

Pies with percentage of flight and/or box plot representing the distribution of PM utilization would ease the comparison.

Chapter 6

What was the gain of considering a vertical profile in the new algorithm?

Line 387-392: These conclusions are quite obvious to me. What are the operational impacts of your results? What does it mean in terms of design? Is there a need to redesign the PM? Should sequencing legs be extended or shortened? What does it mean in terms of TMA/E-TMA delay sharing?... What can be done upstream (e.g., metering?)?

3. Response - round 1

3.1 Response to reviewer 1

Overall, I think the article is pretty much ready for publication, but the (sample) code should be fixed. On the paper, I have the following comments:

Line 120: the sentence with 'granulated' should be changed to something like "...trajectory data in the form of state vectors at 1s resolution."

Response

Thank you for the suggestion. We changed the sentence in the manuscript.

Changes in manuscript: Changed sentence in line 124.

Figure 3: Add a legend for the meaning of shapes (4) and colors (4) [aren't they representing the same concept? If yes, maybe shapes are good enough].

Response

Thank you for your valuable comment. You are correct that the shapes and colors in Figure 3 represent the same concept. To improve clarity, we have added an explanation of the shapes in the legend. However, we believe that using both shapes and colors enhances the figure's visual perception.

Changes in manuscript: Updated legend of Figure 3.

Figure 5: Annotate the plots with the waypoints mentioned in the relevant text, i.e., NONVA, GODLU, BABKU...

Response

The waypoints, together with the circle colors, are mentioned in the corresponding text on lines 182-220. However, this is helpful advice. To enhance the readability of the pictures in Figure 5, we add the names of the respective waypoints.

Changes in manuscript: Added waypoint names to each subfigure in Figure 5.

Lines 210-211: It is not clear how four waypoints (LI739, LI760, LI725, LI748) can be marked by two red circles...again, annotating the plot with the waypoints could help.

Response

Thank you for noticing. Indeed, Figure 5 – (h) is showing only the western Point Merge system with NW and SW sequencing legs, and that is why only two of the four red circles are visible.

Changes in manuscript: Added explanation text on line 219.

Line 270: "We horizontal and" is missing a verb.

Response

This is a very valid comment; we corrected the sentence.

Changes in manuscript: Text on lines 267-270.

Table 3: Add a column with

Table x: Add a table similar to Table 3 but for False Negatives.

Response

Thank you for these suggestions. We updated the table, which now shows the number and percentage of both the false-positive and false-negative flights captured with the enhanced catchment algorithm.

Changes in manuscript: New columns in Table 3.

Table 6: It is not clear what Airport Width, Airport Length, or Rectangle means. My understanding, from lines 299-300, "smallest rectangle around the TMA," is that an axis-aligned geographical bounding box is taken for each airport, then Airport Width (in km) = $\max(\text{lon}) - \min(\text{lon})$, rectangle (km^2) = area of the bonding box, Airport Length (km) = $\max(\text{lat}) - \min(\text{lat})$. Please clarify. (BTW, the smallest rectangle could well be one that is not axis-aligned but rotated.) Lines 289-301: I find this check on correlation with quantities such as airport length, airport width, and airport area quite misfit.

Response

This is a valid comment. The reviewer's understanding is completely correct, but we agree that the explanation provided might have been misleading, so we made an attempt to explain the concept in a better way. Regarding the second part of the comment, we agree that the analysis of correlation of the catchment area circle radius and the relative airport sizes does not provide positive results. However, we believe that it is valuable to report even unsuccessful trials to reduce the possibility of duplication.

Changes in manuscript: Changed text on lines 317-318.

Lines 313-314: Too many "only" instances.

Response

Thank you for noticing; we rephrased the sentence with the aim of reducing redundant words.

Changes in manuscript: Rephrased sentence on lines 331-333.

Lines 328-333: The number in Table 7 does not seem to indicate that the two algorithms catch approximately the same number of flights; the differences are too big! (apart from EIDW East).

Response

We agree that the way the sentence was written was quite misleading and did not correspond to the results shown. We believe that, with few exceptions, the algorithms work similarly for the PM systems with fully overlapping sequencing legs, but further research would be needed to provide any further conclusions.

Changes in manuscript: Update sentences on lines 350 and 352.

Code: Output directory in 'Small example' could be added in the repo (you can just add a "hidden" empty file to have git keep it.)

Response

We think this is a good idea. A folder named "Small_example_output" was added to the Small Example directory, containing the awaited outputs from the codes provided.

Changes in manuscript: Changes in GitHub directory.

Code: I advise against directory or file names containing blanks.

Response

We agree with this advice and plan to name folders and files carefully, following the advice in the future.

Changes in manuscript: -

Code: Spellcheck README.md

Response

This is a valid comment, thank you. We checked and corrected the spelling in both README files: the one in the general "JOAS_Journal_Paper_2024" directory and the one in the "Small example" folder.

Changes in manuscript: Spelling corrected in README files in GitHub repository.

Code: The 06_VP_ENGM.py file fails to run: wrong ident error. Code: The subsequent example scripts are stuck. Code: Script 10 has a syntax error.

Response

That is true, thank you for the information. There indeed was an indentation error, which is now corrected. The subsequent scripts might have been stuck as their input is the output of the previous codes. All the codes were carefully double-checked and should be working correctly by now. We also corrected the syntax error in script 10.

Changes in manuscript: Updated GitHub scripts in Small example folder.

3.2 Response to reviewer 2

In this paper, the authors use a "catchment" algorithm to detect flights flying point merge procedures. This algorithm is based on 2 steps, focusing on the lateral and vertical profile of each flight. The authors also present a sensitivity analysis of some of the parameters used in this algorithm. The paper is well-written and easy to follow, with clear explanations of the proposed methodology. Still, I have some comments:

Regarding the correctness checks, I would like the authors to further elaborate on what they mean by "visual observation". Did the authors just plot all the trajectories and then look for false positives and false negatives manually? How did the authors identify the exact flights that were not identified well by their catchment algorithm?

Response

That is the correct understanding. We plot the arriving aircraft trajectories to check for the false-positive or false-negative flight candidates. In the case there are such candidates, we study them separately and check their horizontal and vertical aspects to make the final decision on their belonging to PM. The detailed identification of the false-negative flight trajectories can be found in Section 4.0.3 Correctness Check – False Negatives in the manuscript. We agree that the description in the manuscript might have been confusing, so we tried to clarify.

Changes in manuscript: Added clarification sentence in lines 267-270.

Regarding Table 7, the authors claim in Line 328 that there are just big differences in Oslo Airport and St. Petersburg Airport. However, in Table 7, EIDW (West), ENBR, ENGM, RKSI and ULLI all have big differences in the number of flights detected between both algorithms. The authors should also acknowledge this fact. Any idea regarding why these big differences are happening? Furthermore, are the flights detected by each algorithm the same flights? The quantity might be similar, but the specific flights detected might be different...

Response

This is a good question! We believe that the differences are caused by the major difference in the algorithms; however, we did not have the opportunity to compare the respective methodologies. The goal of the comparison was to report the differences. Regarding the latter question, we checked the specific flights for a small example (One sequencing leg of Bergen's airport PM system), and the results show that out of 38 flights identified by the traffic library, 27 of them were also identified by our catchment algorithm. Our catchment algorithm identified a total of 95 flights for the same period of time and the same PM system. We think that the difference might be the large amount of flight trajectories going directly from the first PM sequencing leg waypoint to the merge point. To draw any educated conclusions, further analysis would be required.

Changes in manuscript: Added explanation paragraph on lines 353-358.

What are the implications of having the PM usage and utilization rates the authors present in Section 5? Were the PM systems in these airports badly designed? Are the ATC not willing to use these new procedures? Any other reason? I would like the authors to further elaborate on this topic.

Response

This is a valid comment, thank you. The aim of this study is to report on how the current PM designs are used, and we think it could work as a tool for the ATCs to assess whether the current PM design satisfies the needs of the airport and the aims of the design. Regarding the low rates for PM usage and utilization, we think both mentioned reasons might be correct. One possible explanation might be that the PM systems were designed with spare capacity to accommodate the expected increase in air traffic. Another possible reason for implementing PM is to improve vertical efficiency of the flight. The corresponding assessment could be done in the future, comparing the vertical efficiency reported before and after implementation of the PM systems.

Changes in manuscript: Added sentence in 5.2 PM Utilization section in lines 395-398.

4) Miscellaneous/typos:

4.1) Line 142: "unsafe"—I am not sure what the authors mean by this...

4.2) Line 149: "to the point, when the aircraft" — comma not needed

4.3) Line 159: "and if yes, the algorithm proceeds to the second step, we analyze the vertical profiles."

— probably needs a rephrasing: "and, if so, the algorithm proceeds to the second step, in which we analyze the vertical profiles."

4.4) Figure 3: Font size should be increased to improve readability.

4.5) Line 266: The "False negatives" title should also be in bold or italic, like the false positives title.

4.6) Line 270: "We horizontal and.." Please correct this.

Response

Thank you for these comments; we made corrections to the manuscript accordingly.

Changes in manuscript: Corrected typos.

3.3 Response to reviewer 3

Overall: The paper and study are well documented, with numerous figures and reports on a lot of work for data cleaning and for the method to quantify the defined indicators. However, the overall aim and objective of the study are not very clear to me.

What is the purpose of measuring the use of the sequencing legs? Which problem does it solve? How does it contribute to aviation science? What are the operational impacts of the results?

Response

The purpose of this study is to understand how PM systems are currently implemented and to what extent they are utilized in the airports. We conclude that Point Merge systems are usually underutilized, if used at all. We think this study could work as a tool for the ATCs to assess whether the current designs satisfy the needs of the airports and how they could improve arrival performance efficiency. To clarify that, we added an explanation of the aim of this study to the manuscript.

Changes in manuscript: Added sentence on lines 52-54.

No link is made between the utilization of PM and the traffic demand, delay absorption capacity, nor functioning regime (e.g., adapt point merge legs to the traffic demand). Indeed, the number of aircraft does not define the traffic demand (as such, the capacity needed). A notion of time/rate should be added at the entry and exit of the PM systems.

Response

This is a very valuable comment and a great suggestion for future work; thank you! However, as stated before, the aim of this study was to report the current situation, not to investigate the link between the capacity and the PM usage.

Changes in manuscript: -

No link is shown between the various configurations of PM (overlapping leg, one/dual runway) and the results of the utilization.

Response

This is a good suggestion for future research, but for now, we don't see any clear dependency between the PM configurations and the utilization of the results. For example, both Bergen airport and the western PM of Dublin airport both operate PM systems with dissociated sequencing legs, but no clear connection in terms of the PM utilization can be found. To address this comment, we provide a small

summary in the manuscript.

Changes in manuscript: Summary text added on lines 406-412.

What do the enhancements perform, compared to the previous research presented at the 11th Open-Sky symposium, bring in terms of performance? E.g., accuracy, precision, recall F1-score. The same applies for comparison with the algorithm from the traffic library.

Response

The second step of the algorithm—the vertical check—was created to enhance the accuracy of the resulting PM data subsets. The vertical check of the catchment algorithm provides additional identification of the false-positive and false-negative flight trajectories. Regarding the comparison of the performance of our catchment algorithm and the one from the traffic library, the purpose of the comparison is to simply report the differences. We believe that the differences are caused by the major difference in the algorithms; however, we did not have the opportunity to compare the detailed respective methodologies.

Changes in manuscript: Added explanation paragraph on lines 353-358.

Page 1 Abstract: "...the geometry...PM or merging points impact the associated trade-offs." Could you please clarify what the trade-offs you mentioned are?

Response

In the abstract, we refer to the trade-offs described by Eurocontrol in its Point Merge Implementation: A quick guide. 2020. The mentioned trade-offs are between the PM system capacity and its efficiency, which have to be decided while designing the Point Merge system for a specific airport.

Changes in manuscript: We added an explanation to the abstract.

Page 4 In addition to the airport info provided, interesting info to add would be how the point merge is used according to runway use: e.g., In the case of a 1 PM system for 2 runways, is it used in an independent/dependent parallel, staggered approach? The same applies if two systems for 2 runways...In addition to traffic load and demand, the final rate delivered at the airport is also important for calculating the capacity.

Response

This suggestion is very good, and we agree that it would be interesting to see the dependencies. Unfortunately, such detailed information is not available for open access. In future work, contacting Avinor regarding this information might be considered. We also agree that the final rate would be an important and interesting addition to our work, which could be studied in future work.

Table 2: I understood that the number of flights corresponds to the results once data is cleaned/-filtered (removing go around, holdings...). Could you please clarify what the percentage of flights removed from the initial dataset is?

Response

Thank you for this suggestion; we addressed this comment by adding a column representing the percentage of outliers that were removed to obtain the final data subsets.

Changes in manuscript: Updated Table 2.

Line 142: There might be a wording error in that sentence: "This metric indicates the frequency of PM procedure unsafe during the period under consideration." What does "unsafe" mean here?

Response

This is a valid comment, thank you. It was indeed a wording error.

Changes in manuscript: Corrected sentence on line 150.

Chapter 4 on Catchment algorithm: A lot of tuning has been made to capture the relevant flight, including sensitivity analysis for the radius sizes that have to be adapted to each airport. In addition, manual observation is still required to distinguish false positives. If the purpose is to measure the distance flown along the sequencing legs, why not use a direct reference trajectory from an entry to an exit point/or delimited area per flow and per airport/runway and measure the actual distance flown between all tracks and these reference trajectories? The difference should correspond to the additional distance flown along the leg, shouldn't it? As (most of) the flights are direct to the merge point when leaving the legs. Then, as the length of the legs is known, derive the percentage of leg use. To me, using the sequencing leg points is pointless, as, no matter the number of points defining the leg, the measure indicating the maximum delay absorption capacity is the total length of the legs.

Response

Thank you for this comment; we think it is a very interesting suggestion that has the potential to significantly reduce the computational time. However, we think that the suggested method would suffer from the existence of holding patterns or any other deviations from the predefined route trajectories. We focus mainly on the part of the flight trajectory related to the PM sequencing leg. Our argument for using the sequencing leg points is that, according to our findings, the flights rarely use the full length of the PM sequencing legs, which might indicate that the spare capacity of the sequencing legs is unnecessary for the current amount of traffic.

All the figures with tracks and procedures are distorted. Could you please re-do them with the correct projections or orthonormal frame? Here, the equidistances between the sequencing legs and the merge point, which is the basic principle of the PM procedures, are not visible in the figures.

Response

We believe that the figures effectively fulfill their purpose by illustrating the appearance of the trajectories and highlighting the points used in the catchment algorithm.

Chapter 4.0.4 Column titles of table 6 are misleading. PM size is (if I understood well) the length of the sequencing legs. What are Airport width/length and Rectangle? (is it TMA surface?) Some are in NM; others are in km. I would suggest removing it as it does not bring added value, as no correlation is found.

Response

This is a valid comment. To enhance the readability of the table, we adjusted the “PM Size” column name to “PM Length” and added further and careful explanation in the text. Regarding the second part of the comment, we agree that the analysis of correlation of the catchment area circle radius and the relative airport sizes did not yield positive results. However, we believe that it is valuable to report even unsuccessful trials to reduce the possibility of duplication.

Changes in manuscript: Updated column name in Table 6 and clarified text on lines 309 and 311.

Chapter 4.1 Table 7: To enable comparison between the catchment algorithm and the Traffic library, I would appreciate dedicated indicators (accuracy, recall, precision,...) instead of raw difference numbers. At the end, what is the most accurate/efficient algorithm?

Response

We think this is a valid comment, and we agree that the differences between the two algorithms deserve further investigation. To the best of our knowledge, there exists no unified measure of accuracy, as there is not sufficient information about what is the ground truth. We also believe that the high differences might be caused by a loose definition of the PM flights as the flight trajectories performing PM procedures but using 0

Changes in manuscript: Added explanation paragraph on lines 353-358.

Line 324-333: To me, Bergen is not the only airport with fully separated PM. EIDW for runway 10L/R (PM East) could be considered as such (as defined in Table 1).

Response

Thank you for capturing this error; we corrected it in the text.

Changes in manuscript: Corrected sentence on line 346.

How can you explain such a wide variation of performance ranging from -98

Response

This is a valid comment. As stated before, we don’t have a clear explanation for why we observe such results, and we think further investigation would be needed. We also agree that the number of flights does not give a full picture to the analysis. To address this comment, we checked the specific flights for a small example (One sequencing leg of Bergen’s airport PM system), and the results show that out of 38 flights identified by the traffic library, 27 of them were also identified by our catchment algorithm. Our catchment algorithm identified a total of 95 flights for the same period of time and the same PM system. We think that the difference might be the large amount of flight trajectories going directly from the first PM sequencing leg waypoint to the merge point. To draw any educated conclusions, further analysis would be required.

Chapter 5.2 Line 358: Could you please further explain what “PM system not unified” means?

Response

We use the term “not unified” PM system to address the PM systems that have different numbers of waypoints on each of their sequencing legs. The explanation can be found in the manuscript on line 384.

Tables 10, 11...: All these tables are quite cumbersome to read and to compare, even more so with sub-tables according to radius. Why not use only one radius value (revised or initial), e.g., the one that performs better? A Pie chart will ease the visualisation of the radius.

Response

Thank you for this valuable suggestion. We provide results for both the initial and the revised catchment area circle radius because we want to report the improvements in the catchment algorithm. We agree that in future reporting, where we only report the performance, it would be beneficial to provide results only for the best chosen circle size. The catchment algorithm development is a major part of this paper, and we believe that it is valuable to show the difference.

Please make them consistent: ranges/bins of percentage are different for each table: 20

Response

The percentage bins are different for some PM systems, as they don't usually operate with the same number of waypoints on the sequencing legs. In our calculations, we use the percentage based on the last waypoints, the aircraft trajectory visited before turning to the merge point; thus, the percentage bins are defined by the number of the waypoints along the sequencing legs.

At the end, what I'm interested in is, what the PM utilization in terms of distribution: how many go direct (early direct or direct "just at leg entry"), what is the median, quantile, percentile...

Response

We agree that such visualization has the potential to improve the readability of the results. We created pie charts for each airport showing the proportion of non-PM flights, PM flights, and the flights identified as PM but utilizing only the first waypoint (direct to the merge point). We included the pie charts in the manuscript and agree that the visualization enhances the value of the findings.

Changes in manuscript: Added Figure 11 with pie charts and additional text on lines 414-428.

Pies with percentage of flight and/or box plot representing the distribution of PM utilization would ease the comparison.

Response

We believe that Figure 10 – The Cumulative PM Utilization serves the purpose and represents the PM utilization at each of the airports.

Chapter 6 What was the gain of considering a vertical profile in the new algorithm?

Response

As stated above, the second step of the algorithm—the vertical check—was created to enhance the accuracy of the resulting PM datasets. The vertical check of the catchment algorithm provides additional identification of the false-positive and false-negative flight trajectories.

Line 387-392: These conclusions are quite obvious to me. What are the operational impacts of your results? What does it mean in terms of design? Is there a need to redesign the PM? Should sequencing legs be extended or shortened? What does it mean in terms of TMA/E-TMA delay sharing?... What can be done upstream (e.g., metering)?

Response

The purpose of this study is to understand how PM systems are currently operating and to what extent they are utilized in the airports. We conclude that Point Merge systems are usually underutilized, if used at all. We think this study could work as a tool for the ATCs to assess whether the current designs satisfy the needs of the airports and how they could improve arrival performance efficiency. To clarify that, we added an explanation of the aim of this study to the manuscript.

Changes in manuscript: Added explanation on lines 52-54 and 436-438 in the Conclusions.

4. Review - round 2

4.1 Reviewer

Dear authors,

Thanks for having provided answers and some clarifications to my numerous comments with the modifications/addition of text you've made.

Globally, I think further work and analysis is required.

Indeed, some answers to the main comments are, to me, not satisfactory and remain valid:

- There is a chapter on comparison with the traffic library algorithm, but only on numbers of flights caught (not accuracy). With difference values ranging from one extreme to the other without explanations. These differences are analyzed according to the PM design (without a clear link), whereas I think the PM utilization is more dependent on the traffic demand (entry condition) and airport/runway capacity. I find it a pity not to make a link with these key aspects, knowing that the trade-off between system capacity and efficiency is mentioned in the intro.
- It is stated that the addition of a vertical profile enhanced the catching algorithm compared to the previous one presented in the symposium paper, but no concrete figures/numbers are reported on this comparison.
- Finally, the figures of trajectory were not corrected and are still distorted. Plus, in Figure 9, used for comparison, the scales of tracks are different and very small (for the left ones). So, it is, to me, not possible to compare them or spot differences.

5. Response - round 2

5.1 Response to reviewer

There is chapter on comparison with traffic library algorithm but only on numbers of flight caught (not accuracy). With difference values ranging from one extreme to the other without explanations. These differences are analysed according to the PM design (without clear link) whereas I think the PM utilisation is more dependant on the traffic demand (entry condition) and airport/runway capacity. I find really a pity not to make link with these key aspects knowing that the trade off between system capacity and efficiency is mentioned in the intro.

Response

Thank you for this comment; we think it is a good discussion point. But unfortunately, here we cannot report and compare the accuracy numbers, first of all, because there is no ground truth on which of the flight trajectories should actually belong to PM and which not. Each application will have its own

internal definition and the corresponding metric to measure accuracy of the proposed algorithms. In our calculations, we report the number of false positively and false negatively caught trajectories, but again, according to our own definition of the PM flights as described in Section 4.0.3 Correctness Check. Regarding the comparison to the traffic library, to the best of our knowledge, there is no accuracy metric provided.

Regarding the rest of the comment, we want to clarify that the purpose of this proposed work is to present the new metrics for evaluation of the PM procedures and the algorithms describing them, rather than the analysis of how the metrics could be used for further analysis. This is another wide topic for future work. We believe that the designers of each specific PM system at the airport of interest can target different objectives when propose the specific designs. Apart from the capacity considerations, they may want to improve the environmental efficiency of the flights, to reduce the noise, to lower ATCOs workload, etc. Therefore, we think that the dependency on traffic demand and entry conditions alone may not be sufficient to evaluate the full impact of the PM usage. A comprehensive evaluation of the effect of the PM procedures should be addressed in future work.

But to address the comment, we further examined calculated PM usage per-hour and tested it for dependency from the flight's intensity. We tried the average PM usage per hour with the number of flights which resulted in very low correlation (0.46). Similarly, we tested the median values of the PM usage per hour against the number of flights which resulted in moderately strong correlation (0.72). Additionally, we calculated the PM utilization per hour statistics and tested their dependency on the number of flights. The dependency of the number of flights and the PM utilization per hour average values is low (0.56) and similarly the correlation test for the number of flights against the PM utilization per hour median values resulted in moderate correlation (0.63).

It is stated that the addition of vertical profile enhanced the catching algorithm compared to the previous one presented int symposium paper, but no concrete figures/numbers are reported on this comparison.

Response

Thank you for this comment; we will try to clarify the concerns. In the paper, we provide Table 3, which shows the numbers of false-positive and false-negative flights that were identified with the second step of the catchment algorithm, the vertical check. But we agree that the positioning of the table was rather unfortunate; therefore, we changed the position of the table to fit better in the paper. Further information about what the vertical check brings into the picture can be found in Table 8, which shows the number of identified PM flights with the initial algorithm (only horizontal check) and the new algorithm (two-step algorithm), together with the PM usage numbers.

Changes in manuscript: Changed position of Table 3.

Finally, the figures of trajectory were not corrected and still distorted. Plus, in figure 9 used for comparison, scales of tracks are different and very small (for the left ones). So it is to me not possible to compare them nor spot differences.

Response

This is a very valid comment. We agree that the Figures might have been distorted, and thus we took action to correct that. We corrected the problematic figures (d), e), f) in Figure 5 and Figure 7-d). Additionally, we tried to zoom in in the left pictures in Figure 9.

Changes in manuscript: Changed Figure 5 – d),e),f), Figure 7-d), and Figure 9 – a), c)