

EDITORIAL

Reviews and Responses for Testing Applicability of Point Merge Systems for Göteborg Landvetter Airport

Authors: Henrik Hardell, Tatiana Polishchuk, and Lucie Smetanová

Reviewers: Rainer Koelle, Annette Temme

Editor: Manuel Waltert

1. Original paper

The DOI for the original paper is <https://doi.org/10.59490/joas.2024.7899>

2. Review - round 1

2.1 Reviewer 1

This is a decent piece of work studying the possible deployment of point-merge at Landvetter airport.

The enclosed review comments offer some pointers on how to strengthen the argument.

My major criticism—on a high note—is that the figures/tables do not follow the text. The reader is forced to jump back and forth, and sometimes it is extremely difficult to spot the described, observed, or highlighted aspect.

Please also consider highlighting a key finding of your paper in the conclusion: the use of shorter sequencing legs at airports with lower traffic density and the potential to "grow with traffic." That could be a major outcome of your work.

Overview: The paper "Testing Applicability of Point Merge Systems for Göteborg Landvetter Airport" addresses the challenges of implementing point-merge operations at Göteborg Landvetter aerodrome.

The paper is organized as a use-case analysis and builds on a series of related papers published by the co-authors studying point-merge operations and quantifying potential operational performance benefits. The paper revolves around 'characterizing' the operations at Landvetter for a peak day in 2019, defining arrival clusters, and procedural constraints/parameters. This forms the input for modeling an optimized arrival schedule for which the performance benefits are evaluated.

Writing style: The authors have an engaging writing style. A better distribution of the figures/tables to stay in sync with the surrounding text/explanation could help the reader to better grasp the described phenomena.

Section 1: Introduction

The introduction provides a short background on point-merge as an arrival flow metering technique and points out some related previous work and findings. The authors also cross-reference earlier

work, in particular the perceived benefit from their developed assessment method.

All in all, the introduction is very brief and lacks some explanation or motivation for the work reported. As there is “interest to implement PM in several airports in Sweden,” it might be worth sketching out why Göteborg Landvetter is a valuable use-case application (note: in other parts of the paper, the authors motivate and guide the reader through their thinking—which is an excellent style).

Section 2: Performance Evaluation of Current Arrival Operations

The second section establishes the approach and associated data for studying arrival management at Landvetter. Lines 51-55 provide a good explanation for the reader (cf. above missing motivation—this is one example of how to do it!)

Note on terminology/wording: The authors refer multiple times to a “fair and comprehensive evaluation of ... performance” without specifying what makes it “fair.” These sections can perfectly live without the ‘drumbeat’, e.g., using just “evaluation.”

l.57 feels like an afterthought. The text suggests that the data was prepared earlier or as part of another study. But the statement fails to specify on “what” a per-cluster analysis is performed. Do not assume that every reader will study all your other work (e.g., the bespoke Stockholm Arlanda study in [13]).

l.61/62/63 - ditto. Think about providing at least a listing of what you take from the other work. While this might be part of the following paragraphs, the reader is left to speculate whether this is all.

l.64 - I assume the 6 clusters were established in [13] .. help the reader to navigate by providing enough breadcrumbs (or assumptions/given facts) you are using in your work.

l.66ff - it might be easier on the reader to list the measures as a bullet list or table (which might be a cross-referenceable artifact regarding mentioning what you use/take from previous work). I do not want to split hairs—but rethink whether the measures should be “bold” (as the key to take away) rather than italic. But I might fail to see the logic/taxonomy here.

l.86 - cf. above. How about turning this into a 4-5 line paragraph and summarizing “what” you apply from your earlier work (or why).

Section 3: Analysis Per Flow/Cluster

This section talks the reader predominantly through Figure 1 and Figure 2 showing the (cluster of) arrival flows/associated flight-time-to-final heatmaps and the established metrics per cluster. It then moves to describe the vertical efficiency measurement observed for the different arrival flows (i.e., 6 clusters).

For an uninitiated reader, the section’s start is a wild ride: We analyze ... first rows [...] correspond to ... Horizontal Spread ...

I have blind trust in the reported values (and the work), however, it is not trivial to navigate the findings or follow the course of presentation. There are trends that—to be frank—are hard to see in the visualizations. Maybe there is a better way to position the Figures and the surrounding text?

Table 1 – while it is mentioned later in the text, the previous paragraph speaks about specific values and clusters. This illustrates what I want to point out: The text and objects (tables, figures) are not in sync and make it difficult to follow the argument.

l.102/104 - is the similar shape across the different clusters surprising? Why is Cluster 2 “inconsistent”? 500 seconds on final under 10 minutes – what portion of the arrival control takes place during this window? Why is this different from the other (more consistent) shapes/trends (of the distribution)?

Note: I have not had the time to cross-check your earlier work. X secs to final might imply that you are not cutting off at the threshold (but on a gate, e.g., x NM from threshold). This is an important parameter to specify/name in your paper, as it will help to answer some of the questions about procedures/dispersion.

Note: The goal of air traffic services is to sequence flights (orderly and) safe, i.e., safe separation between successive arrivals. How are your 70 or 100 secs showcasing this? Think also about annotating your visualizations, if you point out such specifics (again: guide the reader to look at what you see).

A question that pops up going through this section is how the clusters (and associated traffic patterns in Figure 1) are related to the “sequencing effort” and the higher/lower effort observed? (sorry for sounding like a broken record: what is a sequencing effort value of 183 telling me/the reader [outside studying your earlier work])

Any views on the expressiveness of the results of cluster 4? How is the number of flights playing a role here? Is then your statement about “significantly more time” valid / representative for the cluster meriting a deeper investigation?

l.122 speaks to Figure 3b – with an implicit subsetting of the clusters to the “busiest” ones. Is it useful to combine Figure 3a and 3b in one graphic?

l.126/127 have an important point in terms of “demography” (characteristics) of your dataset. How could this be better introduced / highlighted above?

Maybe l.128/129 could motivate to have a better description of the operational context of Göteborg/-Landvetter earlier in the paper?

Section 4: Point Merge Implementation (? Title vs work reported, how about ‘simulation’)

This section works out the operational characteristics of non-overlapping point-merge arcs applied to Landvetter. The work involves the modeling of point-merge operations and simulating/assessing the suitability and success of the arrival metering.

l.147ff are a good example of how to help the reader follow the logic of your approach by reiterating “what” was taken from previous work (i.e., provide a general overview) and how this is applied to the studied context of Landvetter. Well done!

l.159 - what are the criteria for identifying these “periods” (during which arrival optimization may give the anticipated benefits)?

l.159ff cross-references a variety of additional data sources/sets. Would it be worth aggregating the “materials” in a (sub-)section and augmenting your reported “methodology”? There is also a well-conceptualized ‘trajectory execution model’ at the heart of your experimental work (that deserves a bit more limelight).

l.188 provides reference to a cut-off at IF (initial/intermediate arrival fix ?) – this feels like essential information to highlight the x-secs before threshold cut-off. What would be useful is to have an appreciation of what distance this represents. This gives a pointer to what portion of the time “established on final” is covered and influences the separation on final.

Sections 4.1 and 4.3 are sort of one/two paragraph sub-sections. Could you picture them going elsewhere? Think about better balancing your content / think about whether these are sub-sections on their own right or can be integrated into another section.

Sub-section 4.4 Experimental Evaluation This section details the simulation work and associated (performance) measures. It kicks off with (another) dataset description/specification and develops the optimized arrival schedule. L232 refers to a waypoint W6 (which can be found back in Fig. 4 – if one searches for it). Help the reader to navigate your approach. W6 appears to the entry fix (probably western entry fix). This could be a better label to help someone remember information presented elsewhere.

From sub-section 4.4.2, you reference Fig. 7 and Fig. 8 showing the optimized arrival schedule. The notion of point-merge sequencing is shown in Fig. 8 (with the orange ‘lubber’ line – which is hard to spot and how it relates to the text in 4.4.2). The text reflects the figures. The figure itself is the consequence of your ‘modeling’ / optimization algorithm.

Is there another way to present the ‘optimized’ arrival sequence this text is after (and build the case of using point-merge at Landvetter)?

4.4.3 talks over the observed performance comparing the average total time and distance vs the benefits of lower fuel consumption.

Good development of argument regarding the benefits from the optimized top of descent and CDO profiles.

L291 – can you think about how a ‘fairer’ comparison of certain segments in your clusters could look like? Regarding Table 2 and Fig. 9 you correctly exclude clusters 4 and 5 due to the small sample size of 1 flight each. It might be better to remove these clusters from Fig. 9 and Table 2 or at least grey them out in Table 2. Ditto on Figure 10.

L282 – this is an interesting observation. The wording is a bit irritating. The way you describe it, it feels like the heatmap shows that the “considerable earlier organization” of the arrival traffic coincides with the point-merge legs/procedure airspace.

With these results, would it make sense to showcase the entry of a flight into the procedure vs leaving it (cf. above – what is your optimization result?)

L288-290 – comprises the main conclusion / sales argument and then you specify this in the paragraph L291-300

Section 5 – Conclusions

The concluding section is relatively brief and focuses on the two major building blocks of the paper: i.) analysis of arrival traffic/flows at Göteborg/Landvetter and ii.) (experimental) analysis of the (potential) point-merge operations. Major numerical results are reiterated as a summary statement.

Personally, I think a “half-page” conclusion of a paper of 14 pages is too short “by definition.” You tackled the “what was done/shown” part, however, there is no pointer to the strengths/weaknesses of your approach/method/work, how it can be applied, and possibly future work or ideas.

A key point of your work is a bit weak in the conclusions: if there are benefits for deploying a point-merge design with shorter sequencing legs that can “grow with traffic,” I think you should make sure that the reader finds this back in the conclusions. If this observation can be generalized it could offer concepts like alternative entry points for procedure arcs at shorter legs dependent on traffic density, etc.

English

Congrats, all in all, the paper is well written and has undergone some proofreading. Outside 1.220 som"e," the paper is free from typos. Terms are used consistently throughout the paper. As referenced above, think about the use of terminology and phraseology and/or concepts as such versus trumpeting effects, e.g., "fair" evaluation. From an aviation purist perspective, I prefer English spelling over US. But that will not invalidate the work.

2.2 Reviewer 2

Thank you for sharing your paper "Testing Applicability of Point Merge Systems for Göteborg Landvetter Airport." Your paper is well written and it is easy to follow the main aspects. However, to make it easier to understand the details, I recommend adding some equations either directly or as an appendix. Let me give some examples and more detailed recommendations:

- Although you refer to your relevant earlier literature for most methodologies, I recommend adding the main equations used directly in this paper. Examples where I miss the mathematical definitions are especially the different efficiency calculations applied.
- In the introduction, you describe work using simulated annealing or MIP. Although these are well-known optimization strategies, I'm not sure if you can assume that they are well known by the audience of an engineering journal. Please spend 1-2 sentences to describe these techniques.
- In line 38, you state that you used a clustering technique (already described in an earlier paper). Please describe the kind of clustering technique used and add the optimization equation of the clustering technique and the values of parameters used.
- In line 56, you say that you "curated" a previously known dataset to fit your purpose. What does that mean? Please explain the techniques you applied and why. Could these techniques have influenced the results?
- 1.198/199: Please add more details regarding the optimization model and add the main equations from your earlier literature.
- 1.216: Matlab is well-known but Gurobi needs some explanation and reference.
- Three figures in one row are really small in print, especially for figure 10. The levels you mention for figure 10 (a) are hard to identify visually. Have you assessed how your data curation influences the actual routes and your evaluations for figure 10 (a), especially the identification of levels flown?
- 1.289/290: You state "These results support that the introduction of the PM procedures noticeably improve the metering and spacing efficiency within TMA." Does that hold if you relax the assumption of a simplified standardized arrival weight? Part of the simplified traffic structure visible in figure 10 (b) compared to 10 (a) may be due to eliminated spread in arrival weight.
- A few abbreviations (e.g. MIP, ASMA) are not explained.

3. Response - round 1

3.1 Response to reviewer 1

Section 1: Introduction The introduction provides a short background on point-merge as an arrival flow metering technique and points out some related previous work and findings. The authors also cross reference earlier work, in particular the perceived benefit from their developed assessment method. All in all, the introduction is super brief and misses a bit of explanation/motivation for the work reported. As there is "interest to implement PM in several airports in Sweden" it might be worth to sketch out why Göteborg Landvetter is a valuable use-case application (note: at other

parts of the paper, the authors motivate and guide the reader through their thinking – which is an excellent style).

Response

We have added that the interest includes both Stockholm Arlanda and Göteborg Landvetter. As already stated, we say that ‘this analysis is relevant since there is currently an ongoing internal project on redesigning procedures and TMA for the airport’, as well as it is the second biggest airport in Sweden.

The second section establishes the approach and associated data for studying arrival management at Landvetter. L.51-55 provide a good explanation for the reader (c.f. above missing motivation – this is one example on how to do it!)

Note on terminology/wording: The authors refer multiple times to a “fair and comprehensive evaluation of ... performance” without specifying what makes it “fair”. These sections can perfectly live without the ‘drumbeat’, e.g. using just “evaluation”.

Response

Thank you for your comment. We agree that using more relaxed wording better serves the purpose. By “fair” evaluation, we mean that the framework has already been tested across multiple airports with different configurations, and we believe it can effectively capture performance without favouring any particular features or specifics. To address this comment, we changed wording in the related sentences.

l.57 feels like an afterthought. The text suggests that the data was prepared earlier/as part of another study. But the statement fails to specify on “what” a per-cluster analysis is performed. Do not assume that every reader will study all your other work (e.g. the bespoke Stockholm Arlanda study in [13]).

Response

This is a valid comment; the purpose of pointing out to our previous work is our practice to mitigate redundant methodologies, however in this case it could have been addressed better. We added a short explanation of the clustering technique together with the corresponding equation (1).

l.61/62/63 - ditto. Think about providing at least a listing of what you take from the other work. While this might be part of the following paras, the reader is left to speculate whether this is all.

Response

To address this comment, we added a summary of the metrics used in the paragraph.

l.64 - I assume the 6 clusters were established in [13] .. help the reader to navigate by providing enough breadcrumbs (or assumptions/given facts) you are using in your work.

Response

We agree with this comment and provide a better explanation in the text.

l.66ff - it might be easier on the reader to list the measures as a bullet list or table (which might be a cross-referenceable artifact regarding mentioning what you use/take from previous work). I do not want to split hairs – but rethink whether the measures should be “bold” (as the key to take away) rather than italic. But I might fail to see the logic/taxonomy here.

Response

Thank you for expressing your opinion regarding the organization of the text, we split the block of text into a bullet point list, and we agree that it is now easier to navigate for the reader. However, we decided to keep italic text for the metric names as they are not constituting the main contribution of this work but rather supporting the main purpose of the first part of the paper, the detailed per-flow performance evaluation.

L.86 - c.f. above. How about turning this into a 4-5 line para and summarizing “what” you apply from your earlier work (or why).

Response

To address this comment, we added a couple of sentences briefly explaining the functionality of the optimization framework that was developed earlier, in Section 2.2.

Section 3: Analysis Per Flow/Cluster

This section talks the reader predominantly through Figure 1 and Figure 2 showing the (cluster of) arrival flows/associated flight-time-to-final heatmaps and the established metrics per cluster. It then moves to describe the vertical efficiency measurement observed for the different arrival flows (i.e. 6 clusters).

For an uninitiated reader, the section’s start is a wild ride: We analyze ... first rows [...] correspond to ... Horizontal Spread ...

I have blind trust in the reported values (and the work), however, it is not trivial to navigate the findings or follow the course of presentation. There are trends that – to be frank – are hard to see in the visualizations. Maybe there is a better way to position the Figures and the surrounding text?

Response

We agree that the trends in Minimum Time to Final pictures might be difficult to observe; to help navigate the reader, we add directions to observe the color scale bar. Furthermore, we split original Figure 1 into two smaller Figures (1 and 2) to better fit into the text.

Table 1 – while it is mentioned later in the text, the previous para speaks about specific values and clusters. This illustrates what I want to point out: The text and objects (tables, figures) are not in sync and make it difficult to follow the argument.

Response

This is also a valid comment; we divided original Figure 2 into two Figures (3 and 4) which fit better into the text and the position of Table 1 is now located closer to the affiliated text.

L.102/104 - is the similar shape across the different clusters surprising? Why is Cluster 2 “inconsistent”? 500 seconds on final under 10 minutes – what portion of the arrival control takes place during this window? Why is this different from the other (more consistent) shapes/trends (of the distribution)?

Response

Regarding the reported observed similarity in shape of the Spacing Deviation evolution curves, we believe this is not surprising as all the arrivals in clusters are navigated using the same arrival procedure. In our previous work, we suggested that different shapes and different spreads of the 90th quantile curves differ among airports with different arrival procedures. (H. Hardell, A. Lemetti, T. Polishchuk, L. Smetanova. Evaluation of the Sequencing and Merging Procedures at Three European Airports Using Opensky Data. MDPI proceedings to OpenSky Symposium 2021, November, Brussels.)

We refer to inconsistent spacing in Cluster 2 as the overall deviation from the spacing of two consecutive aircraft (quantile curves) is quite high and suggest that the air traffic controllers experienced difficulties maintaining the given spacing intervals. We agree that this sentence might sound misleading and changed the wording.

The arrival control during this window applies the techniques (vectoring in this case) to space and sequence the aircraft to the final approach. We believe the difference in the shape could be caused by the fact that the arriving aircraft can be navigated through various paths (west and east from the runway) to the final approach fix.

Note: I have not had the time to cross-check your earlier work. X secs to final might imply that you are not cutting off at the threshold (but on a gate, e.g. x NM from threshold). This is an important parameter to specify/name in your paper, as it will help to answer some of the questions about procedures/dispersion.

Response

The trajectories are cut based on the TMA border coordinates. We agree that the Minimum Time to Final heatmaps might be misleading as they appear that the trajectories are cut inside the TMA. However, as can be seen on the real trajectory plots, it is not the case. Parts of trajectories in the heatmaps appear to be missing as the values of the Minimum Time to Final in those cells exceed the maximum limit of the colorbar. We think that shifting the maximum of the colorbar to higher values would jeopardize the visualization of the results closer to the runway.

Note: The goal of air traffic services is to sequence flights (orderly and) safe, i.e. safe separation between successive arrivals. How are your 70 or 100 secs showcasing this?

Response

We attempted to describe it by the phrase “The low dispersion of the curves around 70 seconds to final for all clusters indicates successful sequencing before the final approach”.

Think also about annotating your visualizations, if you point out such specifics (again: guide the reader to look at what you see).

A question that pops up going through this section is how the clusters (and associated traffic patterns in Figure 1) are related to the “sequencing effort” and the higher/lower effort observed? (sorry for sounding like a broken record: what is a sequencing effort value of 183 telling me/the reader [outside studying your earlier work])

Response

Sequencing effort is quite a new metric which, to the best of our knowledge, wasn’t studied extensively in the past and thus no standardized values exist. We think that the sequencing effort has different “normal” or average value per each airport and the observed trends provide insight on what actions

required less or more than usual effort from the air traffic controllers. The sequencing effort is described more in detail in our previous work, where its application on three different European airports can be seen. T. Polishchuk, L. Smetanová. New Insight Towards Characterization of the Terminal Areas. AIAA Aviation Forum 2023.

Any views on the expressiveness of the results of cluster 4? How is the number of flights playing a role here? Is then your statement about “significantly more time” valid / representative for the cluster meriting a deeper investigation?

Response

We believe that more than the number of flights, the geometry and the various trajectory options (west and east from the runway) play its role in the difference of the results for Cluster 2. We observe a higher number of flights in Cluster 3 with significantly lower dispersion of the spacing deviation curves and lower sequencing effort.

L.122 speaks to Figure 3b – with an implicit subsetting of the clusters to the “busiest” ones. Is it useful to combine Figure 3a and 3b in one graphic?

Response

The initial decision to combine Figures 3a and 3b into one graphic was to use the space in the paper in the most effective way. Unfortunately, Figures 1-4 are all quite big and it is difficult to position text around them. To address this comment, we have tried multiple variations of the Figures positioning (splitting, wrapping text around, etc.) but keeping them in one graphic proved to be the most effective in terms of space usage.

L.126/127 have an important point in terms of “demography” (characteristics) of your dataset. How could this be better introduced/highlighted above?

Response

We agree that it would be a good idea to introduce this earlier in the text. We added a few sentences, after Figure 1 has been introduced for the first time, explaining that the airport itself is not located in the center of the TMA, hence, the distance and time aircraft spend inside TMA differ, when the different clusters are compared.

Maybe L.128/129 could motivate to have a better description of the operational context of Göteborg/-Landvetter earlier in the paper?

Response

This was obviously missing, thanks for the suggestion. We added a short description of the operational context in Section 1.

Section 4: Point Merge Implementation (? Title vs work reported, how about ‘simulation’)

Response

Good suggestion; changed to ‘Simulation’.

This section works out the operational characteristics of non-overlapping point-merge arcs applied to Landvetter. The work involves the modeling of point-merge operations and simulating/assessing the suitability and success of the arrival metering.

p.147ff are a good example of how to help the reader follow the logic of your approach by reiterating “what” was taken from previous work (i.e. provide a general overview) and how this is applied to the studied context of Landvetter. Well done!

l.159 - what are the criteria for identifying these “periods” (during which arrival optimization may give the anticipated benefits)?

Response

This sentence is only to explain the process. Later in the text, we explain that we apply our optimization on the busiest day of 2019, which was our goal. The idea was to find a period with sufficiently many flights to be optimized, so that the impact of the optimization can be noticeable.

l.159ff cross-references a variety of additional data sources/sets. Would it be worth to aggregate the “materials” in a (sub-)section and augment your reported “methodology”? There is also a well-conceptualized ‘trajectory execution model’ at the heart of your experimental work (that deserves a bit more limelight).

Response

We agree, this is a good suggestion. We added a short paragraph summarizing the different sources we use and for what, in Section 4.2.

l.188 provide reference to a cut-off at IF (initial/intermediate arrival fix ?) – this feels like an essential information to highlight the x-secs before threshold cut-off. What would be useful it to have an appreciation of what distance this represents. This gives a pointer to what portion of the time “established on final” is covered and influences the separation on final.

Response

The distance from IF to the threshold is 11.25 NM. The IF further connects to the ILS approach; this information has now been added to the text.

Sections 4.1 and 4.3 are sort of one/two paragraph sub-sections. Could you picture them going elsewhere? Think about better balancing your content / think about whether these are sub-sections on their own right or can be integrated into another section.

Response

We agree that 4.1 can be integrated into 4. Since we have added more text to Section 4.3, we think it deserves its own subsection. Hence, we will keep 4.3.

Sub-section 4.4 Experimental Evaluation

This section details the simulation work and associated (performance) measures. It kicks off with (another) dataset description/specification and develops the optimized arrival schedule.

l.232 refers to a waypoint W6 (which can be found back in Fig. 4 – if one searches for it). Help the reader to navigate your approach. W6 appears to the entry fix (probably western entry fix). This could be a better label to help someone remember information presented elsewhere.

Response

A reference back to Figure 4 has been added to guide the reader.

From sub-section 4.4.2, you reference to Fig. 7 and Fig. 8. showing the optimized arrival schedule. The notion of point-merge sequencing is shown in Fig. 8 (with the orange 'lubber' line – which is hard to spot and how it relates to the text in 4.4.2). The text reflects the figures. The figure itself is the consequence of your 'modeling' / optimization algorithm. Is there another way to present the 'optimized' arrival sequence this text is after (and build the case of using point-merge at Landvetter)?

Response

It has been added to the text, where Figure 8 is referred to, that PM usage is identified by the vertical line below the dot indicating the arrival time.

4.4.3 talks over the observed performance comparing the average total time and distance vs the benefits of lower fuel consumption. Good development of argument regarding the benefits from the optimized top of descent and CDO profiles.

l.291 – can you think about how a 'fairer' comparison of certain segments in your clusters could look like?

Response

One obvious solution in order to obtain a fair comparison of certain segments in the clusters is to perform the optimization for an hour during which the distribution of arrivals is better spread over the six clusters, or just find other busy hours where the actual performance of the flights/flights in a certain cluster was different compared to what it is in the selected hour for our paper. Drawing conclusions from one flight from one cluster does not make much sense, since we don't know what the reason of this flight's inefficiency was.

Regarding Table 2 and Fig. 9 you correctly exclude clusters 4 and 5 due to the small sample size of 1 flight each. It might be better to remove these clusters from Fig. 9 and Table 2 or at least grey them out in Table 2. Ditto on Figure 10.

Response

Good suggestion regarding Table 2; clusters 4 and 5 have been greyed out. Results for clusters 4 and 5 have been removed from Figures 9 and 10.

l.282 – this is an interesting observation. The wording is a bit irritating. The way you describe it, it feels like the heatmap shows that the "considerable earlier organization" of the arrival traffic coincides with the point-merge legs/procedure airspace. With these results, would it make sense to showcase the entry of a flight into the procedure vs leaving it (c.f. above – what is your optimization result?)

Response

Exactly! Here we highlight that the heatmap and Spacing Deviation evolution curves nicely visualize the time points when the aircraft sequence becomes better organized and help us to find the exact moment of time when that happens. But from these figures, we cannot see whether aircraft enter or leave the procedures at this point in time.

Section 5 – Conclusions

The concluding section is relatively brief and focuses on the two major building blocks of the paper: i.) analysis of arrival traffic/flows at Göteborg/Landvetter and ii.) (experimental) analysis of the (potential) point-merge operations. Major numerical results are reiterated as a summary statement.

Personally, I think a “half-page” conclusion of a paper of 14 pages is too short “by definition”. You tackled the “what was done/shown” part, however, there is no pointer to the strengths/weaknesses of your approach/method/work, how it can be applied, and possibly future work or ideas.

A key point of your work is a bit weak in the conclusions: if there are benefits for deploying a point-merge design with shorter sequencing legs that can “grow with traffic”, I think you should make sure that the reader finds this back in the conclusions. If this observation can be generalized it could offer concepts like alternative entry points for procedure arcs at shorter legs dependent on traffic density, etc.

Response

We agree that the conclusions section is a bit weak. We have added an explanation about that we cannot draw any final conclusions on what size PM system is suitable, since we would need to perform additional experiments with different traffic situations. We would also evaluate the performance of PM for the opposite runway and make the same simulation as we already did, but with an added time buffer for the separation requirements. We also believe that our tool can be used to evaluate different PM designs for an airport, which we added to this section.

3.2 Response to reviewer 2

Although you refer to your relevant earlier literature for most methodologies, I recommend adding the main equations used directly in this paper. Examples where I miss the mathematical definitions are especially the different efficiency calculations applied.

Response

We agree that this is a valid comment. We added an Appendix with selected equations or pseudocodes for most of the metrics used.

In the introduction you describe work using simulated annealing or MIP. Although these are well-known optimization strategies, I’m not sure if you can assume that they are well-known by the audience of an engineering journal. Please, spend one to two sentences describing these techniques.

Response

Good idea. We added a short description on MIP in Section 4.3.

In line 38 you state that you used a clustering technique (already described in an earlier paper). Please, describe the kind of clustering technique used and add the optimization equation of the clustering technique and the values of parameters used.

Response

Thank you for these comments; we have described the clustering technique in Section 2.1.

In line 56 you say that you “curated” a previously known dataset to fit your purpose. What does that mean? Please, explain the techniques you applied and why. Do these techniques may have influence

on the results?

Response

We understand that the explanation of the choice of the dataset might be misleading. In this work, we reuse a previously created dataset for arrivals to Göteborg Landvetter airport. The dataset was extracted from the Opensky Network and was carefully cleaned with our standard pre-processing procedures: removing outliers, smoothing the trajectories, and applied interpolation when some data was missing. When working with the dataset in this study, we noticed further outliers which we needed to remove from the initial one in order to provide quality performance evaluation.

I.198/199: Please, add more details regarding the optimization model and add the main equations from your earlier literature.

Response

We have added the main equations and some more text explaining the optimization model in Section 4.3.

I.216: Matlab is well-known but Gurobi needs some explanation and reference.

Response

We have added a sentence to explain what Gurobi is and that it is used to solve the MIP, formulated in AMPL.

Three figures in one row are really small in print, especially for figure 10. The levels you mention for figure 10 (a) are hard to identify visually. Do you have assessed how your data curation influences the actual routes and your evaluations for figure 10 (a), especially the identification of levels flown?

Response

We do not expect the reader to identify all the levels in Figures (a) and (b) visually, but suggest that the reader compare these two figures and see that the profiles in (b) are steeper in general and feature longer levels (flat segments). We choose to keep the three figures in one row for space saving. Data curation does not change the shape of the actual vertical profiles or levels flown.

I.289/290: You state “These results support that the introduction of the PM procedures noticeably improve the metering and spacing efficiency within TMA.” Does that hold if you relax the assumption of a simplified standardized arrival weight? Part of the simplified traffic structure visible in figure 10 (b) compared to 10 (a) may be due to eliminated spread in arrival weight.

Response

We assume in this case that by the word arrival weight the reviewer is referring to the weight of the aircraft that we assume. If this is the case, we think it is a valid comment, and testing our optimization framework with more randomly assigned aircraft weights (within a certain range) could be an interesting future work to explore. Aircraft may, in this case, descend differently compared to when using the standardized weight, but our optimization framework would work with any descent profile. Whether it is able to find a feasible solution or not, is another question.

A few abbreviations (e.g. MIP, ASMA) are not explained.

Response

Explanations have now been added where the abbreviations first appear.

4. Review - round 2

Thank you for providing the revised version of your paper. Addressing the comments has greatly improved the understandability of this interesting use case study to introduce point-merge operations at Göteborg Landvetter airport. I recommend accepting this version.