

Busting the Myth! - Reproducibility in ANS Performance

Enrico Spinielli, Rainer Koelle
Performance Review Unit
EUROCONTROL
Brussels, Belgium
enrico.spinielli@eurocontrol.int
rainer.koelle@eurocontrol.int

Abstract—Air transportation is undergoing a fundamental transformation. On the political, strategic, and operational level work is directed to meet the future growth of air traffic. To meet this goal, operational concepts and technological enablers are deployed with the promises of distinct performance improvements. These improvements are typically widely marketed, however, the underlying evidence is not made available. From that respect Air Navigation System Performance suffers from the same credibility crisis seen in other industries. This paper implements the reproducibility paradigm by providing data, code, and results openly to enable another researcher or analyst to reproduce - and potentially - improve or build on our work. This work is part of a wider thread to establish Open Data and Open Source based data analytical capabilities for Air Navigation System Performance in Europe. The implementation of the paradigm is demonstrated on the basis of the performance reference trajectory by replicating a classical performance measure, and then moving into investigating the application of trajectory based information for enhancing the current state of performance monitoring for the arrival phase. This paper applies the concept and approach as an use case analysis for 2 European airports with different strategies for sequencing arrival traffic (i.e. stack holding and point merge). The work reported demonstrates the feasibility of the approach and reflects a stepping stone in the roll-out of the performance reference trajectory within the European context. The work also builds the basis for novel developments on the basis of the implemented reproducibility workflow as this paper is accompanied by a webpage that supports the impartial evaluation of the work by providing the data, code, and how this reports had been produced.

Keywords—*reproducibility; ANS performance; flight trajectory; open data; performance indicator*

I. INTRODUCTION

Air transportation is undergoing a fundamental transformation. To meet the future growth of air traffic, political decision makers, strategic planers, practitioners and researchers aim to enhance the performance of air navigation services (ANS). At the time being there is no consistent and regionally harmonised deployment programme. Typically, the deployment programmes are revolving around public-private partnerships and regional funding programmes resulting in a wide mix of consortia and stakeholder subsets. The actual deployment of novel concepts or technological enablers is then driven by the

programmes and to a lesser extent by operational network wide priorities following a piece meal approach.

Moreover, the success of these deployments is typically supported by promising business cases and operational evaluations - often expressed as operational performance improvements. While these improvements are marketed, the underlying data and associated analyses in support of the results are seldomly made available. Explanations range from data sensitivity, commercial interests and property rights, through to complexity of the data analyses or simulations, and the associated volume of data.

Accordingly, ANS performance improvements are subject to a “credibility crisis” similar to reported problems in other domains, e.g. medicine [1], psychology [2], ecology [3]. Fidler and Wilcox (2018) provides a recent review of the reproducibility discussion [4]. Part of the crisis sensation was driven by the diffusion of reported results as other researchers/analysts were able to discover substantial flaws in the data analyses. However, a crisis also offers room for self-correction. Throughout the recent years, this lead to revisiting the concept of reproducible research. This concept is based on the premises that an independent researcher/analyst is enabled to reproduce the reported results. Reproducibility covers therefore a spectrum and mix of method, data, and analytical code/scripts per se openly available for scrutiny by others.

Independent (operational) ANS performance builds on transparent and robust performance monitoring and reporting. Within the European context EUROCONTROL’s Performance Review Unit (PRU) is committed to provide impartial analyses of the performance of the European Air Navigation System based on verified data. With a view to enhance transparency, PRU is - inter alia - making performance-related data publicly available via its Portal, <http://ansperformance.eu>, since early 2017 combined with the methodological documentation for the calculation of the performance indicators used under the EUROCONTROL Performance Review System and the European Union Performance Scheme. There has been recognition of using common and openly available tools and data in

ATM research and operational validation. With the advent of crowd sourced data and aviation databases there is an opportunity to advance the state of the art by combining open data and government organisation collected data to establish a curated platform supporting the quest for reproducibility in ANS Performance. One of the associated building blocks is the performance reference trajectory [5]. The performance reference trajectory will be deployed as Open Data throughout 2019. The trajectory project builds on crowd sourced surveillance data fused with correlated position reports collected from European air navigation service providers. Based on the Open Data model it will be possible to augment the data processing (e.g. data coverage, time synchronisation), establishment of derived data bases (e.g fleet data base), associated performance models (e.g. aircraft performance model), but also the continual further development and validation of operational performance measures/indicators.

This paper builds on the previous work and implements the reproducibility paradigm. As a use-case example, a fully reproducible paper is devised assessing the performance in the arrival phase for 2 European airports. The underlying data, supporting data transformation and analysis, and the paper “source text” are made openly available at the paper companion web repository <https://github.com/euctrl-pru/reproducible-ans-performance-paper>. This enables the independent review and validation of the work reported. Conceptually, an interested researcher/analyst can reproduce this paper by using the same toolset or validate the results by using the source data or interim data products if the complete reproduction of the study would be too time/resource consuming.

The contributions of this paper are as follows:

- first fully reproducible ANS Performance related paper: The paper is written in RMarkdown allowing for the analysis of how text, code chunks, and visualisations are produced (including associated parameter settings) on the basis of the “source text”.
- transparent data analyses and result generation through supporting web-page: All data used in this paper, intermediate data artifacts, and results including the associated transformation are publicly available for inspection, validation, or reproduction.
- use-case analysis of ANS Performance during the arrival phase for 2 European airports: As an example the classical ‘additional ASMA time’ performance indicator is implemented on the basis of the provided data with showcasing trajectory derived capabilities to further study the efficiency of the arrival phase (e.g. impact of sequencing strategies [stack holdings, point merge]).

II. BACKGROUND - REPRODUCIBILITY

Conceptually, reproducibility is not a new topic. It is as old as the scientific method and at the heart of the verification/validation of results. The consistent and robust interpretation of results by independent researchers/analysts forms the

evidence and the inherent body of knowledge in support of the applied methods or conclusions. This principle is also relevant for the further development on the basis of such results.

The terms replication and reproducibility are often used interchangeably and there is a debate about the necessary and sufficient conditions for these terms. Results are considered replicable, “if there is sufficient information available for independent researchers to make the same [consistent] findings using the same procedures with new data” ([6], p. 3f). Replicability, thus, does not aim to produce the “same” (aka identical) results. Here the goal is to establish confirmation of the approach and findings by deriving same conclusions.¹ Accordingly, results are reproducible, if an independent researcher can recreate the findings based on the given information. In our domain (i.e. operational ANS Performance and associated data analysis) this will require the access to data and derived summaries, how associated visualisations have been generated, the accompanying explanatory text, and conclusions drawn for the analysis.

Reproducibility has also to be seen in light of today’s “publish or perish” context. It appears that higher emphasis is given to the frequency of publication than the validity of the produced results. Within the medical world cases like the Duke ovarian cancer scandal [1] are not single incidents, but also demonstrate how errors during the data analysis - in this case during the data preparation phase using spread sheets - can discredit the work. Despite the increasing awareness, the adaption of reproducibility or moving to other open scholarship/investigation techniques is not yet achieved. This can be contrasted by emerging practices like data journalism. One interesting example is BuzzFeed’s article on the use of surveillance planes operated by law enforcement and military for a variety of purposes. The authors were able to demonstrate based on Open Data that these purposes range from tracking drug trafficking to testing new surveillance technology [7]. The credibility of the article comes from the fact that all artifacts are made available to reproduce the findings or study the chosen approach and data transformations made leading to the conclusions of the article.

Arguably there are limits to the access to underlying source data. However, these limitations also offer an opportunity to address reproducibility [8]. For example, the data volume size might prohibit the further dissemination of data, its collection is too resource intensive, or there exists restrictive license agreements. Experience has shown that even in such cases, a limited sample of the source data can be made available for others to verify the initial data processing steps. For that purpose the creation of an analytical data set as the output of the data collection, gathering, and cleaning process stage can be understood as sufficient in the sense of reproducible

¹For example, the same data collection procedures may not necessarily result in “identical” results as the time horizon might be different, (new or different) data for another study case show different result values, but the conclusions are consistent with the replicated approach.

research. One key transparency aspect is the documentation of the data preparation steps. The article *Spies in the Skies* by Peter Aldhous and Charles Seife [7] is a good example for a reproducible analysis of the US government's airborne surveillance program. Though the license restrictions did not allow for the publication of the underlying aircraft position data, the authors made available the analytical data set and documented their analysis from the data preparation through to the publication stage.

The reproducibility discussion focuses strongly on the recreation of results by other researchers/analysts. A related problem is the re-generating of already completed (and potentially published) results with the - previously - available same input data. This is a fundamental issue that can be summarised as “can a future me re-do what I had done in the past”? This problem space ranges across the complete data analysis lifecycle, including archiving the originally used input data or the steps to generate the study data, documenting the data cleaning and transformation steps including the parameters chosen, etc. By definition such a capability is a precondition for reproduction or replication by others. This implies that the analytical steps including the steps to produce summaries or estimates, visualisations, and associated conclusions are documented and archived. In order to enable others to recreate the own findings, such steps and documentation would need to be shareable. This consideration has a strong bearing on the toolset chosen for this paper.

Some of the reasons why researchers/analysts do not publish data and computational procedures that allow reproduction will be hard to change [9]. The predominant technical reasons have been listed above. However, there are also personal reasons such as the fear for floating ideas or loss of competitive advantage. Other reasons deal with more technical problems, e.g. lack of capabilities, procedures.

Another incentive for the adoption of an open and reproducible research practice can be seen in the wider recognition through increased citations, media attention, discussion with the community and collaborators, and ultimately funding opportunities. The recent years have seen a strong use of research blogs and community platforms supporting the publication of smaller steps or intermediate results. The combination of web presentation and discussions offers a good basis for an initial dialogue. Platforms like google scholar or researchgate.net offer additional features for outreach and communication. This equally applies for paper archives such as arxiv.org for pre-prints. Based on the credibility crisis discussion a number of journals are now promoting - or even requiring - the provision of analytical data in support of the paper publication. There is also an increasing number of requirements for national or European funded research with respect to data management and Open Data access to the research.

Reproducibility is discussed in various domains. The following examples further promote the use of open source software and/or data:

- archaeology: Marwick et al. (2017) focusses on computational reproducibility in archaeological research by providing basic principles. The case study of their implementation uses open source software (R/Rmarkdown/git approach) [10]. The paper advocates the publishing (analytical) data in support of the paper. This should not be an issue, even if there are constraints on the overall study data.
- engineering / signal processing: Vandewalle (2009) discuss the motivation and approach to reproducibility in signal processing (i.e. what, why, and how). Also here the case is made to open access to the code and data for validation purposes [11].
- operational ANS Performance monitoring: Spinielli et al. (2018) highlight the development and use of Open Flight Trajectories for reproducible ANS [5].

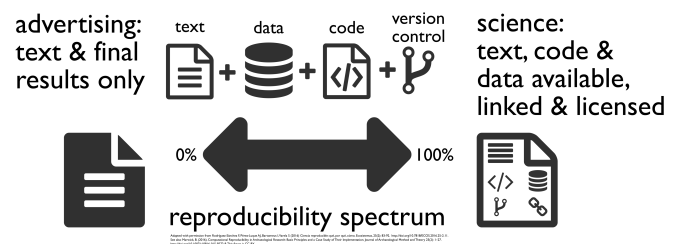


Fig. 1. Reproducible research, Fig. 2 in [10].

Across the literature, researchers have identified a “reproducibility spectrum” ranging from the extreme, i.e. publication only (advertising and marketing of the done work), to almost or even fully replicable research. The work of Marwick et al. (2017) [10] and Peng (2011) [12] uses the analogy of a spectrum that can be mapped against the scoring of Vandewalle (2009) describing the level of reproducibility on a 0 to 5 scale [10]. In light of [13], this paper conceptualises reproducibility as the “ability to recompute data analytic results given an observed dataset and knowledge of the data analysis pipeline” and accordingly “replicability of a study is the chance that an independent experiment targeting the same scientific question will produce a consistent result”.

With Fig. 1 and the discussion in this section, this paper implements a fully reproducible workflow. This is achieved by establishing this paper with open source software which allows to combine text, code, and visualisations. The associated source text can also easily be read with a simple text editor/viewer. From that perspective, the text is fully inspectable. This includes the majority of visualisations which are generated based on the data. Accordingly, the parameters and settings chosen can be validated by an independent researcher/analyst. Last but not least, the data for this paper is made available in conjunction with this text as Open Data. This also includes intermediate data artifacts which allows for the reproduction of the complete work or portions of it. The conceptual approach and its building blocks will be presented

in more detail in the next section.

III. METHOD / APPROACH

A. Conceptual Approach

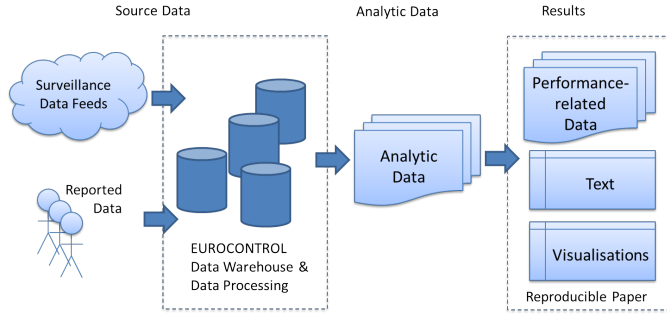


Fig. 2. Conceptual approach to reproducibility by PRU.

This paper builds on the conceptual approach presented in Fig. 2 to implement replicable data analyses for performance monitoring and associated research. On the data level this approach abstracts three levels of data:

- source data
- analytical data
- results (and performance related data)

The source data for this paper comprises the PRU Performance Reference Trajectory and operator reported data under the airport operator data flow. This study also uses some aeronautical information for the use case airports accessible through AIP or community maintained open data archives (e.g. <http://openaip.net>, <http://ourairports.com>). These data flows are processed and stored in the EUROCONTROL data Warehouse. The approach recognises that the full source data may or cannot be provided for a variety of reasons. These include volume of the source data sets, technical constraints allowing access to the source data storage, and data sharing policies or use limitations. For the latter cases, PRU promotes the sharing of a so-called representative sample. The idea is that with the provision of the associated processing code, other researchers/practitioners are able to reconstruct a part of the analytic data - and subsequently part of the reported results - on the basis of the provided source data set.

Analytical data refers to the data sets established based on the aforementioned processing code. An essential defining characteristic of analytic data is that it is the data basis on which the reported results can be produced. The abstraction between source data and analytic data allows for the sharing of such data as limitations are typically only applicable to the source data. We follow the notion of [10] by referring to the provision of a representative sample of 8 days accompanying this paper as source data.

As depicted in Fig. 1 the provision of the data, code, and text allows to move to the right part of the reproducibility spectrum, away from marketing and final results to credible,

transparent, and challengeable analysis. Accordingly, the results (or performance related data) level presents the final abstraction.

B. Toolset

This work builds on the RStudio tools for the R language [14] including Git (and the web-based repository managers GitHub and GitLab) as underlying version control system. The R language was originally developed within the statistical community supporting the task of statistical reporting by providing routines for the statistical computing and visualisation. Being open source, the R community is actively engaging and sharing related software packages to enhance the core functionality. Without limiting the impact of other packages, the development of knitr [15] and RMarkdown [16], ggplot [17] for visualisation, and the so-called tidyverse packages and RStudio IDE [18] provide the basis for an open source data analysis ecosystem.

A key feature for the implementation of the reproducibility paradigm is the fact that RMarkdown supports the combination of text, coding, and visualisations in a single document. The latter can be rendered into various output formats, e.g. classical paper report-styles (pdf, MS Word), html web-pages, and slides (both classical Powerpoint style or html-based presentations). The fact that outputs can be rendered to html pages also supports the embedding of interactive graphics. This also facilitates the communication and outreach amongst researchers/analysts as standard web technology becomes a standard outlet format. It follows that changes to the code (and associated visualisations) or text can be readily updated by re-rendering the adapted report. This allows to develop standard templates for recurring reports based on varying input data, but it also serves the iterative development of larger reports (and studies).

A minimal example of RMarkdown document, a plain-text file with the conventional extension `.Rmd` is shown below:

```

---
title: "R Markdown example"
author: "John Doe"
date: "2019-02-17"
output: html_document
---

Some math

$$\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$
,
followed by numbered list:

1. italics and bold text
1. then bullets
   - simple, eh?

and a code chunk:

```{r}
library(ggplot2)
fit = lm(mpg ~ wt, data = mtcars)
b = coef(fit)
ggplot(mtcars, aes(x = wt, y = mpg)) +
 geom_point() +
 geom_smooth(method = lm, se = TRUE)
```

```

The slope of the regression is ``r b[1]``.
```

# R Markdown example

John Doe

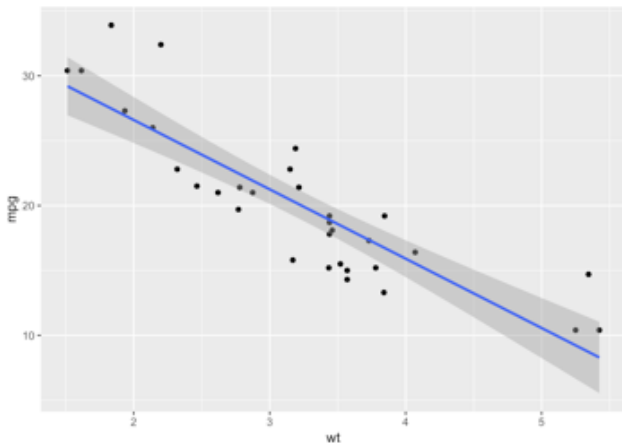
2019-02-17

Some math  $\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$ , followed by numbered list:

1. *italics* and **bold** text
2. then bullets
  - simple, eh?

and a code chunk:

```
library(ggplot2)
fit = lm(mpg ~ wt, data = mtcars)
b = coef(fit)
ggplot(mtcars, aes(x = wt, y = mpg)) +
 geom_point() +
 geom_smooth(method = lm, se = TRUE)
```



The slope of the regression is -5.3444716.

Fig. 3. The output of a minimalist example of RMarkdown.

There are three components of an R Markdown document: the metadata, the text and the code. The metadata, written between the pair of three dashes ---, helps in the production of output document (title, type of document [web page, PDF, MS Word], ...) The text, which follows the metadata, is written using Markdown, a simplified formatting syntax. There are two types of (computer) code:

- A code chunk starts with three backticks and `r` indicates the language name,<sup>2</sup> and ends with three backticks.
- An inline R code expression starts with backtick `r` and ends with a backtick.

Code portions are executed when the document is rendered.

<sup>2</sup>R Markdown supports 50 other languages.

The result of rendering the R Markdown above is shown in Fig. 3.

A particular strong point is that the R/Rstudio ecosystem is open source software. This includes also its extensibility with other packages. There exists a wide variety of packages for data analytical tasks and techniques. This mechanism can also be useful for the communication of results. Researchers/analysts are able to provide their code and supporting documentation as a package for other researchers to build on. For example, throughout the last year, PRU has developed certain utility functions for reoccurring tasks of the data analytical pipeline. It is now planned to wrap these functions into a package saving other researchers/analysts of recoding certain steps. For this paper, the functions and data transformations are reported as part of the RMarkdown and supporting R files. All these files are available via the project webpage.

## C. Capabilities: Performance Reference Trajectory

Throughout the recent year PRU developed the data analytical capabilities to fuse correlated position reports and ADSB surveillance data. This work has been reported in [8] and [5]. The idea is to establish a pan-European air situation picture openly available to research and practitioners. Conceptually, the fused performance reference trajectory will be a key milestone in combatting marketing myths. For example, operational improvements can be directly assessed by interested analysts with a view to validate the observable performance improvements. This allows for the application of commonly accepted methods and models to express performance benefits and - ultimately - challenge related publications on deployments and achieved improvements.

End of 2018 the development project was closed and PRU is now investigating options for deploying and maintaining the processing modules on a cloud platform. The goal is to establish and operate the underlying data processing capabilities throughout 2019 and launch the regular provision of the performance reference trajectory as Open Data. It is also planned to establish a network and complement the air situation picture by deploying additional sensors or determining non-ADSB derived positions as a community effort. As a mid-term goal associated data bases and models for performance shall be established and jointly maintained.

## IV. RESULTS

### A. Data Preparation

The raw data sources for the use cases of this paper have been processed by a pipeline of R scripts. For example, the additional ASMA time indicator and the anticipated further breakdown of the arrival phase can be described by a set of transition points:

- segment I: entry into the ASMA area to entry into the holding/vectoring segment

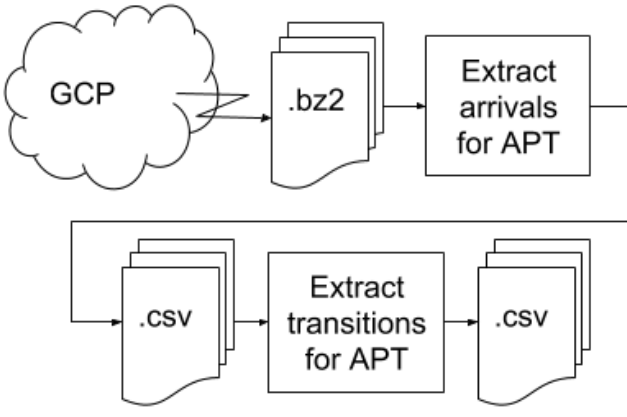


Fig. 4. Data preparation pipeline.

- segment II: holding/vectoring segment; for the use case reported holding stacks or point merge procedure
- segment III: final approach alignment and approach

Figure 4 shows how the transition points (40 NM, entry/exit from holding/point-merge area, landing) for the use case airports have been extracted. Due to the size of the daily reference trajectory files, the pipeline operates on one date at the time. Assuming the relevant RT flight and position files are downloaded from GCP as described in [19], the steps and the relevant scripts (available in the companion repository for this paper) used are:

1. the script `R/extract-arrivals.R` extracts arrivals to the APT airport and generates:
  - a file with flight details
  - a file with all the arrivals' positions
  - a file with all arrivals' positions within 40 NM the airport's reference position (ARP)
2. the script `R/extract-transition-points.R` extracts transition points in the terminal area and generates a file containing per each flight:
  - the first point within 40 NM,  $P_{40}$
  - the first and last point within the holding/point-merge area,  $P_{HOLD}$
  - an estimated landing point,  $P_{LAND}$

This stage makes use of data extracted/extrapolated from the relevant AIP, see for example Fig. 5, such as ARP, runway name and threshold position/elevation, polygons for holding stacks or point-merge areas. Specifically the polygons have been heuristically defined and enlarged by 2NM in order to cater for the spread of real flown trajectories, see the scripts `R/egll-data.R` and `R/eidw-data.R`.

Each position mentioned above consist of at least

- flight ID
- 4D position (timestamp  $T$ , longitude, latitude, altitude)
- flown distance  $D$

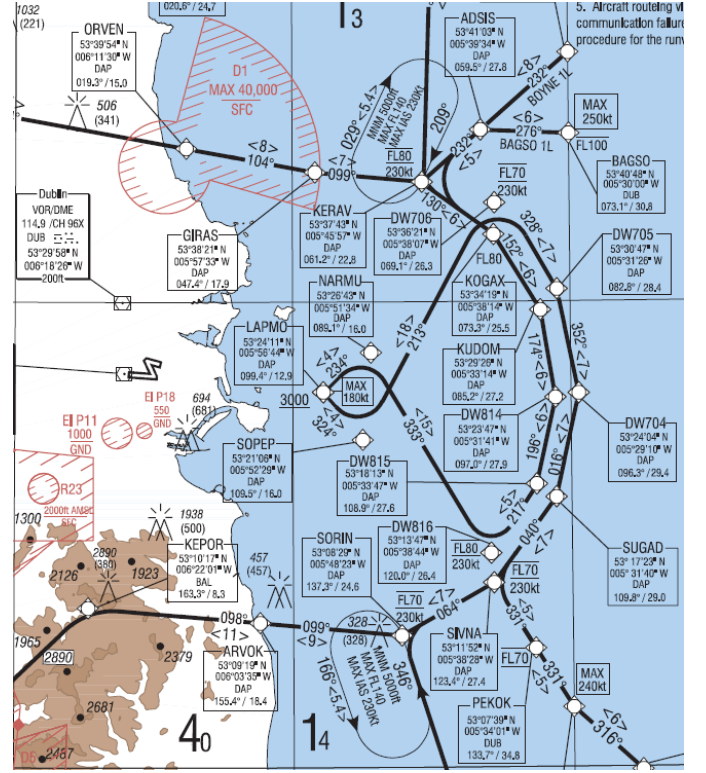


Fig. 5. Excerpt of Irish AIP showing point merge arc layout East of Dublin Airport.

- sequence number
- distance from ARP

These allow to calculate for each flight

- the time spent before the holding/point-merge area,  $T_{PHOLD|first} - T_{P40}$
- the distance flown before the holding/point-merge area,  $D_{PHOLD|first} - D_{P40}$
- the time spent within the holding/point-merge area,  $T_{PHOLD|last} - T_{PHOLD|first}$
- the distance flown within the holding/point-merge area,  $D_{PHOLD|last} - D_{PHOLD|first}$
- the time spent out of the holding/point-merge area till landing,  $T_{PLAND} - T_{PHOLD|last}$
- the distance flown out of the holding/point-merge area till landing,  $D_{PLAND} - D_{PHOLD|last}$

## B. Data Coverage

The previous sections described the data preparatory action. For this paper, a reproducible data set for the analysis of arrival management techniques at London Heathrow and Dublin airport have been extracted. The data set comprises flights for the period of 1.-8. August 2017. It can be inferred from the traffic mix at London Heathrow (major international hub) that only a limited subset of flights is not equipped with ADSB, while a slightly higher level of non-equipped flights is expected at Dublin (higher share of intra-European flights).



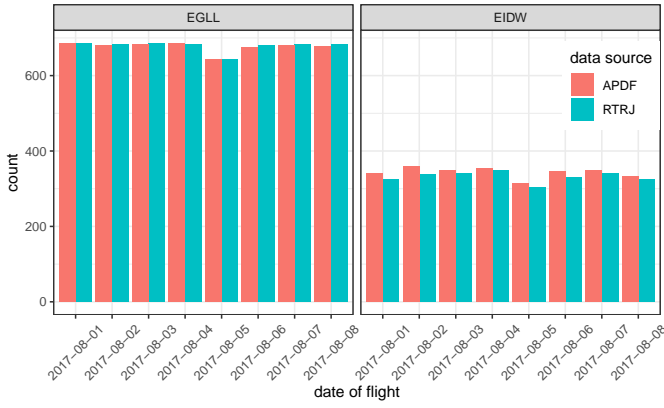


Fig. 6. Number of arrivals in data sets, i.e. 1.-8. July 2017.

Fig. 6 shows the overlap of the data between the arrivals extracted from the airport operator data flow (APDF) and the reference trajectory (RTRJ) for London Heathrow (EGLL) and Dublin (EIDW).

For the time horizon of this paper, the coverage of the reference trajectory derived data is for Heathrow (EGLL) 100% and for Dublin 97%.

### C. “replicate” ASMA (time-based)

The additional time in the arrival sequencing and metering area (ASMA) is a globally recognised measure to assess flight efficiency during the arrival phase. The ASMA metric calculates the additional time based on a reference time which is determined for a family of flights sharing similar characteristics. These are the entry sector into the arrival airspace, weight turbulence category and engine type, and the actual landing runway. For each flight  $i$  the additional ASMA time is

$$add. ASMA time_i = actual travel time_i - reference time_i$$

The reference time  $i$  is determined for flights with similar arrival characteristics (c.f. [http://ansperformance.eu/references/definition/additional\\_asma\\_time.html](http://ansperformance.eu/references/definition/additional_asma_time.html)). These include 1.) the arrival entry sector at 40NM from the airport, 2.) the aircraft weight turbulence category and engine type, and 3.) the actual landing runway. The reference times are built for a longer time horizon and extracted from the PRU data base as additional source data.

As an example, the distribution of the actual arrival transit times for arrivals into Heathrow (EGLL) are shown in Fig. 7. Heathrow operates holding stacks to tactically optimise the arrival throughput. Dependent on the traffic situation, aircraft may experience longer holding times (i.e. long tails of boxplots in Fig. 7).

Fig. 8 compares the additional ASMA time extracted from the PRU performance dashboard, the trajectory based calcu-

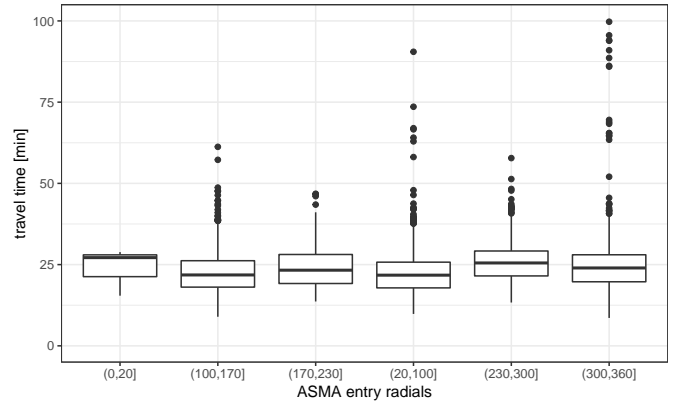


Fig. 7. Distribution of actual ASMA travel times for EGLL (1.-8. July 2017).

lation with the dashboard derived reference time, and a 20th percentile approach purely based on the trajectory data.

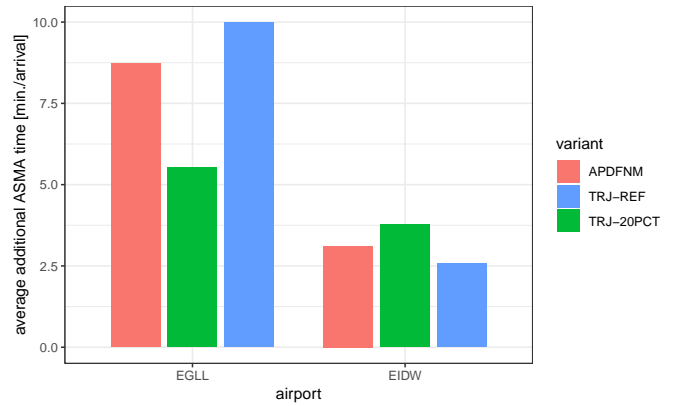


Fig. 8. Comparison of determined ASMA values, i.e. 1.-8. July 2017.

Fig. 8 shows the variation of the timebased additional ASMA times determined for our sample. APDFNM refers to the variant used for the EUROCONTROL Performance Review System. The reference time is build on the basis of a one-year sammple. Applying this reference time to the trajectory based travel times (i.e. TRJ-REF), there is a strong increase of the additional ASMA time for Heathrow (EGLL), while in the case of Dublin (EIDW) reasonable decrease takes place. Calculating the additional ASMA time exclusively with the trajectory sample and applying a 20th percentile approach (TRJ-20PCT), shows a significant drop in EGLL and a moderate increase at Dublin. The high variation for Heathrow shows that the additional ASMA time is strongly dependent on the time horizon. The pure trajectory bases measure - by definition - includes the effect of the operational concept (stack holdings) and distorts the measure. In the case of Dublin it can be argued that the TRJ-20PCT measure reflects the time based ASMA measure sufficiently.

TABLE I  
AVERAGE ADDITIONAL DISTANCE FLOWN (15TH PCT)

AIRPORT	TOT_ADD_DIST	N_ARRS	AVG_ADD_DIST
EGLL	117037.79	5249	22.30
EIDW	33079.22	2634	12.56

#### D. Distance-based ASMA

conceptually, measuring flight inefficiency in terms of flight time accounts for the airspace user perspective. The previous section has also shown the high dependency of the time based measure on the sample size or reference time attribution. Flight time is linked to engine time and accordingly, assuming maneuvering for operational separation and sequencing reasons, fuel burn. From a service provision perspective, efficiency relates to the capability to provide the least distance flown within the ASMA area. In that respect, the data analytical capability to determine an associated additional distance flown metric addresses an open gap.

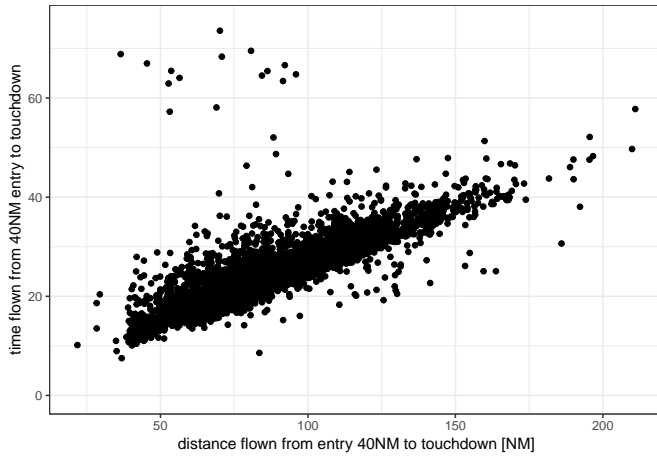


Fig. 9. Relationship travel time and distance flown, i.e. 1.-8. July 2017.

Fig. 9 shows the relationship between the actual distance flown and the actual time flown. The variation of travel times is influenced by varying airspeeds or weather phenomena (e.g. wind). This paper applies a simple 15th percentile approach to the distance flown per arrival sector, aircraft type/engine, and landing runway.

Tab. I lists the determined average additional distance at the study airports.

#### E. Arrival Management Practices

For an initial drill down, this paper uses arrival practices at London Heathrow and Dublin. The approach to managing the arrival flow differs significantly between both airports. Conceptually, London Heathrow employs the “pressure on the runway” concept with a focus on achieving and balancing a continuously high level of runway throughput utilisation. This is achieved by controlling the demand through 4 arrival stacks that allow for the tactical sequencing on the last portion of

the approach. Dublin is one of the airports that pioneered the implementation of “point merge”. Point merge aims at replacing circular holdings by longitudinal holdings to establish the sequence based on predefined legs.

From a performance perspective it is therefore interesting to identify and measure the impact of the operational concept on the overall flight efficiency during the arrival phase of flight.

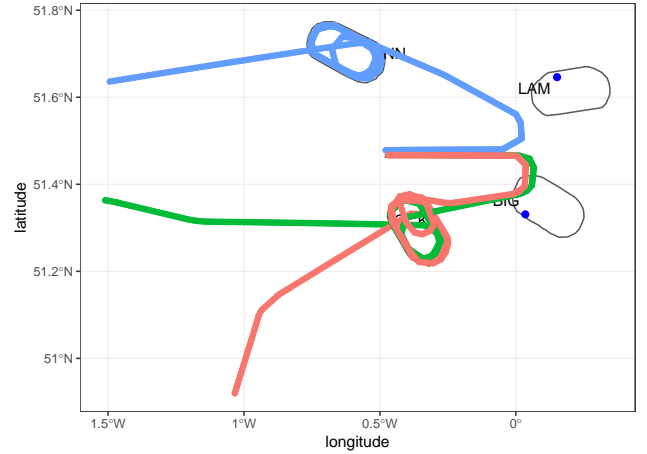


Fig. 10. EGLL - long trajectory flown on 1. July 2017.

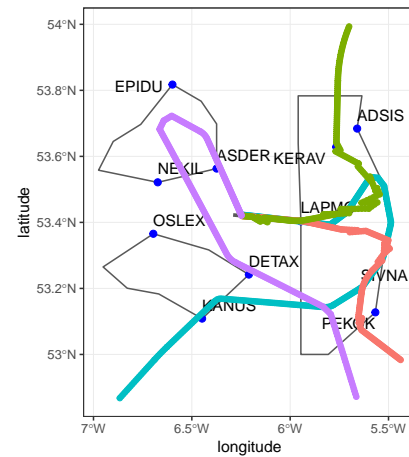


Fig. 11. EIDW - long trajectory flown on 1. July 2017.

Fig. 10 and Fig. 11 show examples for long distances flown on 1. July 2017. The trajectory data allows for the identification of the holdings and/or sequencing area. This allows to separate the effect of the operational concepts.



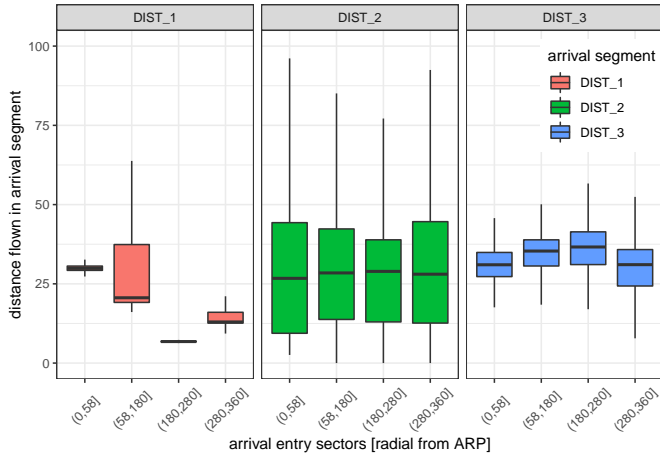


Fig. 12. Actual distances flown for each segment of the arrival phase (i.e. EGLL, 1.-8. July 2017).

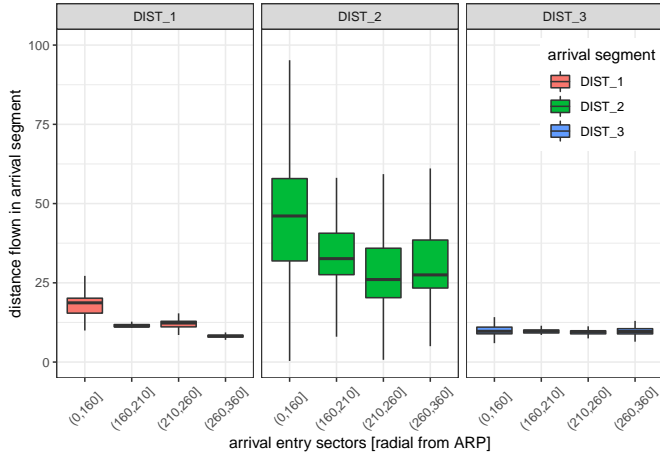


Fig. 13. Actual distances flown for each segment of the arrival phase (i.e. EIDW, 1.-8. July 2017).

Fig. 12 depicts the flown distances per segment of the arrival phase. In the case of London Heathrow the classical stack based holding can be seen in the significant variations of the actual distance flown per arrival direction. Only limited sequencing takes place during the first segment (arrival within 40 NM and the stacks). Fig. 13 shows different footprint. It follows from the distribution of the initial approach segment that this portion is already used for sequencing and synchronisation by Dublin approach. The middle segment shows a wider variation and less uniform distribution than the approach at London Heathrow.

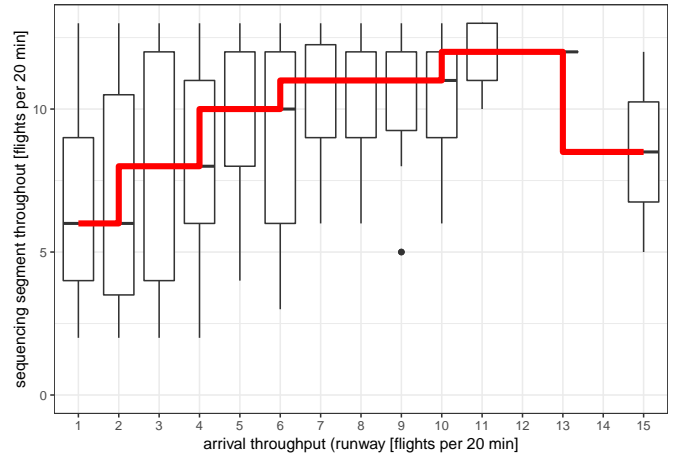


Fig. 14. EIDW - arrival throughput saturation (1.-8. July 2017).

Fig. 14 depicts the variation of the arrival segment throughput at Dublin (EIDW). The median for every set of flights in the ASMA area versus flights landing is highlighted. The behaviour is a classical saturation curve, i.e. up to an arrival bunching of 10-11 flights per 20 min (30-33 flights per hour), the point merge procedure seems to be able to support the tactical absorption and sequencing. It is noteworthy that the saturation curve falls for higher arrival throughputs. As no measurements for 12 through 14 landings had been made in the sample, this behaviour could be caused by the chosen time horizon.

## V. CONCLUSIONS

The work reported establishes a fully reproducible paper for ANS performance in Europe. It is the one of the first attempts to address the reproducibility crisis within ANS performance by providing the artifacts that allow an independent researcher or professional to reproduce the results and verify upstream processing steps.

This paper follows up on previous work [5], [8], [20] and forms part of the roll-out of the performance reference trajectory by the PRU. As such the paper demonstrates the feasibility to the approach, the data analytical capabilities and preprocessing, and the reporting via the R/RStudio ecosystem. This represents an essential milestone moving towards the envisaged target environment. However, further work is required to establish the reproducible environment for ANS Performance.

To overcome these limitations, PRU will be looking into deploying the trajectory processing on a cloud-based platform including the regular provision of the identified analytical data levels.

The move to open data for ANS Performance is a *giant leap* in terms of transparency. This will foster a new level of cooperation and discussion amongst researchers and professionals interested in this topic. It removes the burden to acquire and preprocess large-scale surveillance data and associated aircraft

data or aeronautical information. PRU envisages a strong collaboration through a network of research and practitioners. The departure point for this development was to address the lack and hurdle of verifying results. A further advantage is that novel applications, methods and algorithms (e.g. machine learning), can be applied to a harmonised dataset. It is hoped that development efforts can accelerate while offering a platform for improvements to existing processing methods and algorithms.

The major application of the paper results will inform the performance reference trajectory roll-out project. This goes in hand with the iterative move of the operational performance measures to trajectory derived data. This can take on the form of quality assurance measures (e.g. verifying the order of magnitude, complementing missing reported data), but also making the reference trajectory the primary means of data collection. The latter will support the further development of the operational performance measures responding to the ICAO GANP.

This paper documents a key milestone in the project. As such the future work will address:

- roll-out and establishment of the performance reference trajectory: This includes the organisational and technical work to deploy the processing modules and arrange for a regular processing of the source data and its provision as Open Data.
- establishing the current performance reporting based on the trajectory: Since more than 2 years PRU addresses the need for Open Data by providing monthly updates for the key performance indicators used under the EUROCONTROL Performance Review System and the SES Performance Scheme. Underlying analytic data is made available via <http://ansperformance.eu>. As part of the roll-out, it is foreseen to move associated indicators to a trajectory derived or at least verified process and expand on the analytical data level.
- establishment of a network: The underlying idea is that various levels of future work can be orchestrated in close collaboration with interested researchers and professionals. This can span work addressing the improvement of the actual data processing (e.g. less resource intensive computing) to the further development and validation of novel techniques for measuring operational performance (e.g. application of machine learning, time series analysis).

## REFERENCES

- [1] G. Kolata, "How a New Hope in Cancer Fell Apart," *The New York Times: Health*, 07-Jul-2011.
- [2] S. J. Ritchie, R. Wiseman, and C. C. French, "Failing the future: Three unsuccessful attempts to replicate Bem's 'Retroactive Facilitation of Recall' Effect," *PloS one*, vol. 7, no. 3, p. e33423, 2012.
- [3] H. Fraser, T. Parker, S. Nakagawa, A. Barnett, and F. Fidler, "Questionable research practices in ecology and evolution," *PloS one*, vol. 13, no. 7, p. e0200303, 2018.
- [4] F. Fidler and J. Wilcox, "Reproducibility of Scientific Results," in *The Stanford Encyclopedia of Philosophy*, Winter 2018., E. N. Zalta, Ed.

Metaphysics Research Lab, Stanford University, 2018.

- [5] E. Spinielli, R. Koelle, K. Barker, and N. Korbey, "Open Flight Trajectories for Reproducible ANS Performance Review," presented at the SESAR Innovation Days 2018, 2018, p. 8.
- [6] C. Gandrud, *Reproducible Research with R and R Studio, Second Edition (Chapman & Hall/CRC The R Series)*. Routledge, 2015.
- [7] C. S. Peter Aldhous, "Spies in the Skies," *See Maps Showing Where FBI Planes Are Watching From Above*, 06-Apr-2016. [Online]. Available: <https://www.buzzfeed.com/peteraldhous/spies-in-the-skies>.
- [8] R. Koelle, "Open Source Software and Crowd Sourced Data for Operational Performance Analysis," presented at the USA/Europe ATM R&D Seminar, 2017.
- [9] D. Nüst, M. Schutzzeichel, and M. Konkol, "Opening Reproducible Research: A research project website and blog," Nov. 2018.
- [10] B. Marwick, "Open Science in Archaeology," *Open Science Framework*, Jan. 2017.
- [11] P. Vandewalle, J. Kovacevic, and M. Vetterli, "Reproducible research in signal processing," *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 37–47, May 2009.
- [12] R. D. Peng, "Reproducible Research in Computational Science," *Science*, vol. 334, no. 6060, pp. 1226–1227, 2011.
- [13] J. T. Leek and R. D. Peng, "Opinion: Reproducible research can still be wrong: Adopting a prevention approach," *Proc Natl Acad Sci USA*, vol. 112, no. 6, p. 1645, Feb. 2015.
- [14] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2018.
- [15] Y. Xie, *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. 2018.
- [16] Y. Xie, J. Allaire, and G. Golemund, *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman and Hall/CRC, 2018.
- [17] H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- [18] RStudio Team, *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc., 2015.
- [19] Performance Review Unit, "Reference trajectories for Performance Review of European ANS," 2019. [Online]. Available: <https://doi.org/10.5281/zenodo.2566844>.
- [20] E. Spinielli, R. Koelle, M. Zanin, and S. Belkoura, "Initial Implementation of Reference Trajectories for Performance Review," *SESAR Innovation Days*, no. 7, p. 8, 2017.

## AUTHORS' BIOGRAPHIES

**Enrico Spinielli** is the Data Manager of the Performance Review Unit at EUROCONTROL, Brussels. He holds a Master Degree in Electronic Engineering from the University of Pisa, Italy. His intellectual interests include data analytics, data visualization and map projections.

**Rainer Koelle** heads the Safety and Quality of Service section of the Performance Review Unit at EUROCONTROL, Brussels. He holds a PhD from the University of Lancaster, 2013, in Communication Systems. His research interests include data analytics for operational performance analysis.