# New perspectives for air transport performance

A. Cook and G. Tanner
Department of Planning and Transport
University of Westminster
London, United Kingdom

S. Cristóbal and M. Zanin
The Innaxis Foundation & Research Institute
Madrid, Spain

***Abstract*— The average delays of flights and passengers are not the same. The air transport industry is lacking passenger-centric metrics; its reporting is flight-centric. We report on the first European network simulation model with explicit passenger itineraries and full delay cost estimations. Trade-offs in performance are assessed using passenger-centric and flight-centric metrics, under a range of novel flight and passenger prioritisation scenarios. The need for passenger-centric metrics is established. Delay propagation is characterised under the scenarios using, *inter alia*, Granger causality techniques.**

***Keywords–delay propagation; passenger-centric; metric; flight prioritisation; Granger causality***

*Foreword*—This work is co-financed by EUROCONTROL acting on behalf of the SESAR Joint Undertaking (SJU) and the EUROPEAN Union as part of Workpackage E in the SESAR Programme. Opinions expressed in this work reflect the authors' views only. EUROCONTROL and/or the SJU shall not be considered liable for them or for any use that may be made of the information contained herein.

## I. INTRODUCTION

The average delays of (delayed) flights and passengers are not the same. The air transport industry is lacking passenger-centric metrics; its reporting is flight-centric. Trade-offs between these metrics need to be better understood, as they are observed to move in opposite directions under certain types of flight prioritisation. With growing political emphasis in Europe on service delivery to the passenger, and passenger mobility, how are we to measure the effectiveness of passenger-driven performance initiatives in air transport if we do not have the corresponding set of passenger-oriented metrics and understand the associated trade-offs in the context of delay propagation?

In the 'POEM' (Passenger-Oriented Enhanced Metrics) SESAR Workpackage E project, we have built a European network simulation model with explicit passenger itineraries and full delay cost estimations. A baseline traffic day in September 2010 was selected as a busy day in a busy month – without evidence of exceptional delays, strikes or adverse weather. We compare the effects of novel flight and passenger prioritisation scenarios on new passenger-centric and flight-centric metrics, which assess not only delay but also a range of costs associated with delay. The propagation of delay through the network is also investigated, using complexity science techniques to complement classical metrics.

TABLE I. PRIORITISATION SCENARIOS

| *Type, level* | *Designator* | *Summary description* |
|---|---|---|
| No-scenario, 0 | $S_0$ | No-scenario baselines (reproduce historical operations for baseline traffic day) |
| ANSP, 1 | $N_1$ | Prioritisation of inbound flights based on simple passenger numbers |
| ANSP, 2 | $N_2$ | Inbound flights arriving more than 15 minutes late are prioritised based on the number of onward flights delayed by inbound connecting passengers |
| AO, 1 | $A_1$ | Wait times and associated departure slots are estimated on a cost minimisation basis, with longer wait times potentially forced during periods of heavy ATFM delay |
| AO, 2 | $A_2$ | Departure times *and* arrival sequences based on delay costs – $A_1$ is implemented *and* flights are independently arrival-managed based on delay cost |
| Policy, 1 | $P_1$ | Passengers are reaccommodated based on prioritisation by final arrival delay, instead of by ticket type, but preserving interlining hierarchies |
| Policy, 2 | $P_2$ | Passengers are reaccommodated based on prioritisation by final arrival delay, regardless of ticket type, and also relaxing all interlining hierarchies |

Table I summarises the prioritisation scenarios investigated. They were designed in parallel with the new metrics. For convenience, they are broadly classified according to the agency of the instigating stakeholder. For example, only airlines are currently likely to be able to estimate their own delay cost data in $A_1$ and $A_2$. The policy-driven scenarios $P_1$ and $P_2$ are bolder than the current scope of European regulations. It is essential to explore the context of the model and the metrics in terms of future developments such as Airport Collaborative Decision Making (A-CDM) and, regarding flight prioritisation, the User Driven Prioritisation Process (UDPP). These technical contexts, in addition to the evolving socio-political landscape, are discussed in Section III. This includes a review of the European Union's underpinning regulatory instrument for air passenger compensation and assistance (Regulation 261, [1]), of high-level political objectives, of the Single European Sky performance scheme, and of recent ATM delay performance. A full discussion of the design of our metrics has recently been published [2], whereby a complementary approach is proposed to the understanding of network performance. This is reflected in the cross-section of results presented in Section IV. We turn first to a review of the start of the art.

## II. OVERVIEW OF POEM MODEL AND EXISTING MODELS

### *A. Existing modelling – the state of the art*

Using large data sets for passenger bookings and flight operations from a major US airline, it has been shown [3] that passenger-centric metrics are superior to flight-based metrics for assessing passenger delays, primarily because the latter do not take account of replanned itineraries of passengers disrupted due to flight-leg cancellations and missed connections. For August 2000, the average passenger delay (across all passengers) was estimated as 25.6 minutes, i.e. 1.7 times greater than the average flight leg delay of 15.4 minutes.

TABLE II. PREDICTED PAX TRIP DELAY BY PERFORMANCE CHANGES

| ***Performance change*** | ***Predicted pax trip delay change*** |
|---|---|
| 15-minute reduction in flight delay | -24% |
| Improved airline cooperation policy in re-booking disrupted passengers | -12% |
| Flights cancelled earlier in the day | -10% |
| Decreasing load factor to 70% | -8% |

Source: [5].

Based on a model using 2005 US data for flights between the 35 busiest airports, [4] concurs that "flight delay data is a poor proxy for measuring passenger trip delays". For passengers (on single-segment routes) and flights, delayed alike by more than 15 minutes, the ratio of the separate delay metrics was estimated at 1.6. Furthermore, heavily skewed distributions of passenger trip delay demonstrated that a small proportion of passengers experienced heavy delays, which was not apparent from flight-based performance metrics ([5], [6]).

Using US historical flight segment data from 2000 to 2006 to build a passenger flow simulation model to predict passenger trip times, [5] cites flight delay, load factors, cancellation (time), airline cooperation policy and flight times as the most significant factors affecting total passenger trip delay in the system (see Table II).

An "inherent flaw in the design of the passenger transportation service" has been pointed out [7], in that service delivery to the passenger did not improve in 2008 in the US, despite the downturn in traffic. One in four US passengers experienced trip disruption (due either to delayed, cancelled or diverted flights, or due to denied boarding). Recovery mechanisms in place for disrupted passengers, such as transfer to alternative flights or re-routing, require seat capacity reserves. However, the airline industry wishes to maximise economies of scale, optimise yield management, maximise load factors, and (thus) to minimise seat capacity reserves. In 2008, as airlines reduced frequencies to match passenger demand, higher load factors severely reduced such reserves [7].

Analysing US flight data for 2007 between 309 airports to estimate passenger-centric delay metrics showed [6] that the average trip delay for passengers over all flights was 24 minutes, whilst for passengers on flights delayed by at least fifteen minutes, the average delay was 56 minutes.

Flight-centric and passenger-centric metrics have also been examined [8] by comparing different rationing rules in a model US ground delay programme rationing rule simulator, exploring the trade-off between flight and passenger delay, and also between airline and passenger equity. (We shall return to these results later.)

Turning to more recent work, [9] presents a closed-form, aggregate model for estimating passenger trip reliability metrics from flight delay data from US system-wide simulations. Metrics were derived from the probabilities of delayed flights and network structure parameters. A particularly appealing finding was that the average trip delay of disrupted passengers varies as the square of the probability of a delayed flight and linearly with respect to rebooking delays.

An analytical queuing and network decomposition model – Approximate Network Delays (AND) – studied [10] delay propagation for a network comprising the 34 busiest airports in the US and 19 of the busiest airports in Europe. The model treats airports as a set of interconnected individual queuing systems. Due to its analytical queuing engine, it does not require multiple runs (as simulations do) to estimate its performance metrics and can evaluate the impacts of scenarios and policy alternatives.

Covering 305 US airports in 2010, an agent-based model reproduced [11] empirically observed delay propagation patterns. Estimated passenger and crew connectivities were identified as the most relevant factors driving delay propagation. The probability of such connections were modelled as proportional to flight connectivity levels at each airport.

Almost no current models use explicit passenger data, although this is planned for the AND model (*ibid*.). Also, actual passenger transfer numbers have been used in numerical simulations of a major US hub, where it was demonstrated [12] that each metric studied – terminal transit times of passengers, aircraft taxi times and gate conflict durations – outperformed observed values through the use of a balancing objective function. (As part of our work in SESAR Workpackage E, we are also preparing publications focused on actual transfer passengers at a major European hub.)

## *B. The POEM model – an overview*

POEM models the busiest 199 European Civil Aviation Conference (ECAC) airports in 2010, having identified [13] that these airports accounted for 97% of passengers and 93% of movements in that year. Routes between the main airports of the EU 27 states and airports outside the EU 27 have been used as a proxy for determining the major flows between the ECAC area and the rest of the world. This process allowed the selection of 50 non-ECAC airports for inclusion of their passenger data.

The two principal datasets used to prepare the input data for the model were IATA's PaxIS passenger itineraries and EUROCONTROL's PRISME traffic data. Extensive data cleaning of the source traffic data was required, especially with regard to unreliable taxi-out data and scheduled times, missing taxi-in data and aircraft characteristics (including registration sequencing).

There are approximately 30 000 flights in each day's traffic and around 2.5 million passengers distributed among 150 000 distinct passenger routings. The assignment of passengers to individual flights, with full itineraries and calibrated load factors, was a fundamental component of POEM. All the allocated connections were viable with respect to airline schedules and published minimum connecting times.

POEM is a full gate-to-gate model with passenger connectivities explicitly modelled. Each simulated process is governed by one or more rules [13]. For example, Rule 33 governs realistic decision-making for missed passenger connections due to delays and cancellations (such as dynamic passenger reaccommodation onto aircraft with free seats, using detailed fleet and load factor data) and integrates with the tail-tracked aircraft wait and turnaround (recovery) rules. Two airline case studies, including on-site visits and workshops, focused on developing and testing specific aspects of the model rules in an operational context.

Core cost estimations in the model are with respect to delay costs to the airline, since it is these that drive airline behaviour. The model represents a normative day and the simulation results thus reflect schedule robustness (e.g. with respect to passenger reaccommodation). Using a cloud-computing platform, each full day's simulation took approximately two minutes. As a stochastic model, statistically stable results were produced typically after ten runs (although the results presented are based on fifty runs).

## III. SOCIO-POLITICAL AND TECHNICAL CONTEXTS

### A. *Socio-political context – the passenger imperative*

SESAR's 'Performance Target' [14] refers frequently to the concept of society and the passenger. The 'societal outcome' cluster of KPAs[1], is defined as being of "high visibility", since the effects are of a political nature and are even visible to those who do not use the air transport system. The 'operational performance' cluster[2] is also specifically acknowledged as impacting passengers.

Social and political priorities in Europe are now shifting in further favour of the passenger, as evidenced by high-level position documents such as 'Flightpath 2050' [15] and the European Commission's 2011 White Paper ('Roadmap to a Single European Transport Area', [16]).

However, it has been accepted that there are currently several problems with regard to the implementation and scope of Regulation 261. A roadmap for the revision of the Regulation was published in late 2011 [17]. After various consultations, a memo was released in 2013 [18] detailing key proposed changes, which could become law by 2015, subject to approval by member states. In summary, the key changes are to: (i) initiate passengers' right to care and assistance after two hours of delay, regardless of the length of the flight; (ii) require an airline to re-route passengers onto other carriers (already much commoner in the US) if it cannot re-route onto its own services within 12 hours; (iii) offer passengers the same rights for delays relating specifically to connecting flights, and to extend such rights to compensation for long delays (including arrival delay) caused by any reason; (iv) introduce new obligations (currently none exist) regarding information on delayed or cancelled flights; and, (v) better define 'extraordinary circumstances' that exempt carriers from paying passenger compensation (although proposed changes to the compensation rights will make these more complex, allowing the carriers more time to avoid cancelling flights, for example).

The baseline scenario ($S_0$) rules of the POEM model reflect airline costs typically imposed by Regulation 261 and common practice regarding care and rebooking during disruption [13]. Under the $P_1$ and $P_2$ scenarios, current constraints on airline practice are successively relaxed and the impacts are examined, as presented in Section IV.

### B. *ATM delay performance and model alignment*

TABLE III. KEY PERFORMANCE STATISTICS FOR 2010 AND 2102

| *Metric* | *2010* | *2012* |
|---|---|---|
| IFR flights (million) | 9.5 | 9.6 |
| Total pax (million, EU 27) | 777 | 734 |
| Average dep. delay (mins) | 14.8 | 9.5 |
| Arrival delays > 15 mins | 24.2% | 16.7% |
| Reactionary delays | 46.7% | 45.5% |

Sources: [19], [20], [21].

Table III compares key statistics for 2010 (the year from which the POEM model's baseline day was taken) and 2012 (the latest year for which such statistics were available at the time of press). It is to be noted that the traffic and passenger numbers are similar. Passenger numbers depend on coverage: whereas data from Eurostat [21] describe a small fall between these periods, EUROCONTROL [20] reports an increase. The Eurostat data refer to the 27 Member States of the European Union as of 1 January 2007 (i.e. they exclude Croatia, which acceded in July 2013) and report total scheduled and non-scheduled passengers carried (passengers are not multiply-counted for each stage of a stopping flight or during transfers). The EUROCONTROL coverage extends beyond these EU 27 states, thus including strong growth in Turkey.

Whilst 2010 suffered from a high number of cancellations (due to the Eyjafjallajökull ash cloud in April and May, strikes in France and Spain, and bad winter weather), this had a limited effect on punctuality *per se* [19]. Nevertheless, punctuality in 2010 was at its worst since 2001, even with traffic below 2007 levels after modest growth on the previous year [19]. The average departure delay values include all flights, with delays counted from the first minute and early departures counted as zero delay. The percentage of arrival delays greater than 15 minutes in 2012 reached an all-time low of 16.7% – the changes in punctuality were largely driven by improvements in en-route ATFM delays [20].

The average departure delay for September 2010, the month from which POEM's baseline was selected, was 13.9

[1] Environment, safety, security.

[2] Capacity, cost effectiveness, efficiency, predictability, flexibility.

minutes, and the average arrival delay was 13.6 minutes. As we have detailed more fully [13], the model was calibrated partly using these values, with $S_0$ (baseline) averages of 13.8 and 13.5 minutes, respectively. With similar passenger and traffic volumes already between the two years, the model could also be recalibrated, if required, to reflect the better delay performance of 2012.

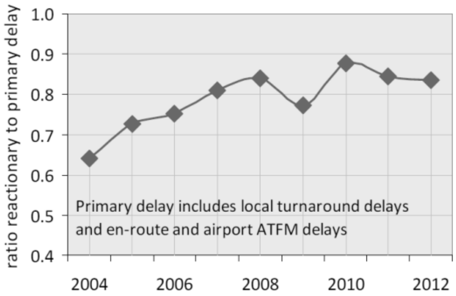


Figure 1. Reactionary delay trend to 2012.
Source: adapted from [20].

Fig. 1 shows the sensitivity of the network to primary delay. In 2010, the ratio was approximately 0.9. On average, every minute of primary delay thus resulted in approximately 0.9 minutes of reactionary delay. After peaking in 2010, the ratio improved in 2011 and 2012. Reactionary delay in September 2010 averaged 46%, with the POEM model $S_0$ value calibrated at 49%.

A key advance made possible through the POEM model is the detailed analysis of the effects of the various scenarios on reactionary delays and the associated trade-offs with other metrics, as will be illustrated in Section IV. The persistence of reactionary delays illustrated through Fig. 1 underlines the continuing importance of research into these effects.

### *C. The SES performance scheme and model flexibility*

EUROCONTROL is designated the Performance Review Body (PRB) of the Single European Sky (SES). The performance scheme is managed by the PRB and is a central element of the SES initiative. It is defined across various reference periods (RPs), as shown in Table IV. Performance targets are set at various levels before each period.

The targets are legally binding for EU member states. These are designed to encourage ANSPs to be more efficient and responsive to traffic demand, whilst maintaining adequate safety levels. The implementation of the performance scheme includes performance monitoring and corrective actions. RP1 addresses mainly the en-route part of air navigation service provision, focusing target setting on en-route capacity, environment and cost efficiency. With regard to monitoring RP1, the PRB has developed an online ‘e-dashboard’ covering all PIs/KPIs[3] at EU, Performance Plan (either national or FAB) and airport levels.

TABLE IV. SES PERFORMANCE SCHEME REFERENCE PERIODS

| *Reference period* | *Applicable years* |
|---|---|
| RP1 | 2012 - 2014 |
| RP2 | 2015 - 2019 |
| RP3 | 2020 - 2024 |

Current ATM key performance indicators (KPIs) in Europe are (inevitably) rather high-level. Many targets have been set [22] at the European level. For capacity (under RP1), ATFM en-route delay per flight (with a weather delay allowance managed at the network level) has a binding EU-wide target of 0.5 minutes by 2014. An intermediate (non-binding) target of 0.7 minutes per flight was set for the first year, 2012, for which reporting was published in September 2013 [23]. The EU-wide value for 2012 was 0.63 minutes of ATFM delay per flight, thus satisfying the target [23]. In fact, en-route ATFM delays decreased by 46% (compared with 2011), albeit with a corresponding 2.7% traffic decrease [23]. Some targets are also applied at the state / FAB level (e.g. targets set on all ‘performance scheme’ airports for total ATFM delay attributable to airport / terminal air navigation services, which take account of severe weather and exceptional events).

RP2 sets out to extend the performance scheme to cover the full gate-to-gate scope, with target setting for four of the International Civil Aviation Organization’s eleven KPAs: capacity, environment, cost efficiency and safety [24]. Union-wide targets for RP2 have to be set by the end of 2013, with adoption of the corresponding Performance Plans in 2014. After a range of consultations and consideration of, *inter alia*, macro-economic conditions, historic performance trends, air traffic forecasts, opportunities for further improvement in ANS performance and associated risks, and evidence of best practice, the PRB published [25] its proposed EU-wide targets for RP2. For capacity, this is 0.5 minutes per flight for 2015-2019. According to PRB analysis [25], this target corresponds to more than 98% of flights not constrained by ATC. (Not a focus of our research to date, the POEM model does not yet have sufficient fidelity for assessing en-route ATFM delays *per se*, although this module is a target for future refinement.)

Setting challenging targets for 2020, SESAR’s Performance Target [14] significantly refines (see Table V) the fifteen minute historical threshold for defining arrival and departure delay in Europe and the US.

TABLE V. SESAR PERFORMANCE OBJECTIVES AND TARGETS

| *SESAR metric* | *Target for 2020* |
|---|---|
| departure punctuality | ≥ 98% of flights departing as planned ±3 mins<br>other 2%: average delay ≤ 10 mins |
| arrival punctuality | > 95% of flights arrival delay ≤ 3 mins<br>other 5%: average delay < 10 mins |
| reactionary delay | 50% reduction by 2020, *cf.* 2010 |
| cancellations | 50% reduction by 2020, *cf.* 2010 |
| variation in block-to-block times | block-to-block σ < 1.5% of route mean[a] |

a. For repeatedly flown routes using aircraft with comparable performance.

[3] KPIs are subject to target setting; PIs are for monitoring only.

Whilst the SES performance scheme focuses on improving air navigation service (ANS) provision, and hence uses ATFM delay in its capacity KPAs, the SESAR targets are broader in scope. Airline punctuality is a poor metric for assessing ANS performance *per se*, since such punctuality is driven to a considerable extent by airline scheduling decisions. Such punctuality metrics remain pertinent in terms of service delivery to the passenger, however, and it is clear that a complementary set of metrics is needed by the industry. Whilst evidence [6] suggests that delays of less than 15 minutes are less important in terms of passenger connectivities, increasing pressures on utilisation and lower connecting times add to the importance of more exacting targets.

For the POEM model results, we focus in Section IV on the trade-offs between flight-centric and passenger-centric metrics, including costs, reporting on the corresponding reactionary delay effects at the disaggregate level, in addition to the impact on the high-level target of Table V. Exploring the robustness of our scenarios under disruption, cancellation rates have currently been used as a model *input* variable, rather than being modelled intrinsically. To measure the effect of increased perturbation, two disrupted days were derived from the baseline traffic. This allowed like-for-like comparisons between the disrupted days and the baseline day. One disrupted day imposed 1 extra minute on the average departure delay (making a new average of 14.9 minutes across all flights). The other disrupted day imposed just under 1% of additional cancellations on morning operations. Comparing the model outputs for the disrupted days showed them to be well modelled in that changes to the core metrics were as expected and reflected operational experience (e.g. with regard to relatively low impacts on flight *punctuality* metrics during periods of higher cancellations). Compared with the baseline day, the scenarios also performed similarly under disruption, demonstrating a degree of robustness in terms of their efficacy under such perturbation (see [13] for details). In 2014, traffic is expected to increase by 2.8%, finally reaching the 2008 pre-economic crisis levels again by 2016 [20]. Future traffic samples could also be used as inputs into the POEM model, which would be interesting to further stress-test the scenarios. (Explicit passenger assignments would have to rebuilt using the dedicated algorithms). It would be feasible, and instructive, to observe the impacts on modelled performance compared with some of the SES / SESAR targets.

### *D. Flight prioritisation and SESAR ConOps*

At the core of the POEM model simulations are the flight and passenger prioritisation scenarios. These need to be considered in the context of the SESAR Concept of Operations (henceforth 'ConOps'). Is there a future role for such mechanisms? If so, over what timescale and at what level of prominence? The SESAR ConOps is mapped into three steps, as shown in Table VI. The three steps are not purely sequential, but overlap, with some changes being developed in parallel across them. The first editions of ConOps Step 1 and Step 2 were published in May 2012 [26] and July 2013 [27], respectively. (ConOps Step 3 will probably be published in 2016.) These documents serve as the main references for all envisaged operationally-related SESAR tasks. Key components of these steps are UDPP and the process into which it feeds – Demand and Capacity Balancing (DCB).

TABLE VI. CONOPS STEPS AND UDPP IMPLEMENTATION

| *ConOps step and operations type* | *UDPP implementation status* | *Other example implementations* |
|---|---|---|
| **Step 1.** Time-based | Initial UDPP; flexibility at dep. airport (reordering flights before pre-departure sequencing, with consistency between airport slots and flight plans) and flexibility en-route (through enhanced ATFM slot swapping) | A-CDM; Network Operations Planning; CTAs (Controlled Times of Arrival); initial SWIM; airport surface management integrated with AMAN/DMAN (arr./dep. manager); some free routes |
| **Step 2.** Trajectory-based | More flexibility to re-arrange schedules in a trajectory-based environment - integrated with en-route CDM; also full integration of AMAN, DMAN and surface management linked to UDPP and (dynamic) DCB | Network Operations Planning based around full 4D trajectories; multiple CTOs (Controlled Times Over) /CTAs (including on non-published waypoints) |
| **Step 3.** Performance-based | Full SWIM (System-Wide Information Management) and collaboratively planned network operations with UDPP | Two airspace categories (civil and military); specific separation tasks delegated to flight deck; free routes implemented from TMA exit to entry; Dynamic Mobile Areas |

Primary sources: [27] and [28].

UDPP is a CDM-based process carried out for DCB purposes, which allows airlines to request a priority order for flights affected by restrictions arising from unexpected capacity reductions. The desired priority order is that which "best respects the business interests" [27] of the airspace users. As explained in ConOps Step 1 [26], (dynamic) DCB with short-term ATFCM measures "constitute a step forward to close the gap between ATFCM and ATC", with the objective of anticipating and managing traffic peaks, and smoothing ATC workload through the application of fine-tuned measures.

ConOps Step 1 extends the previous scope of UDPP. Its deployment phase is from 2014 to 2025. Previously, the emphasis of UDPP was on implementation after DCB had failed to reach an acceptable solution. Its current scope, however, embraces strategic, pre-tactical and tactical phases and will be available in any 'normal' situation, although with a particular applicability during capacity constraints with an early focus, once the design has sufficiently matured, on the pre-departure phase (but ultimately including en-route and arrival phases). Indeed, in the second edition of the ATM Master Plan [28], the prominence of UDPP in the implementation of Step 3 is clear: ""Performance-based Operations" is realised through the achievement of SWIM and collaboratively planned network operations with User Driven Prioritisation Processes (UDPP)." If AO iterations do not come to an agreed solution, the Network Arbitration function makes sure that in the UDPP process a resolution for the conflict between trajectories demand and constraints due to network capacity is proposed. As a last resort, the Network Arbitration function decides a solution. Arbitration follows pre-agreed rules, and is monitored by the Network Manager in its role as UDPP referee.

Step 1 starts from the ‘Deployment Baseline’, which comprises operational and technical solutions that have successfully completed the R&D phase and have already been implemented, or are being implemented, and runs up to 2018. Enhanced ATFCM and DCB processes are part of the Deployment Baseline. A-CDM and UDPP are major milestones for airport integration with the network. From the Deployment Baseline onwards, A-CDM increases information-sharing between the airport, airlines, handling agents, ANSP and the Network Manager to improve cooperation and increase predictability in order to optimise resource utilisation. Indeed, in the airport context, UDPP in Step 1 was developed only for airports operating A-CDM [27]. During steps 2 and 3 (with deployment after around 2025), exchanges of positions in flight lists managed by either the ACC, Network FMPs or the DCB actors at airports will be implemented, from flight planning to the execution phase, and will form the basis of AO negotiation.

Clearly, there is a well-defined place for flight prioritisation strategies within the SESAR ConOps. Indeed, its scope has recently been extended [26]. Already aligned with A-CDM implementation plans, UDPP is a perfect vehicle for the inclusion of cost- and passenger-focused prioritisation mechanisms. In the next section, we demonstrate how the impacts of the POEM (flight) prioritisation scenarios are reflected through appropriate metrics and analytical tools.

## IV. KEY RESULTS FROM THE POEM MODEL

### *A. New metric results*

Fig. 2 presents the core results across various flight-centric and passenger-centric metrics, by the various scenarios. The values indicated[4] are scenario values minus the corresponding baseline ($S_0$) value. Flight prioritisation scenarios ($N_1$ and $N_2$) operating during arrival management based simply on the numbers either of inbound passengers or on those with connecting onward flights, were ineffective in improving performance. The policy-driven scenario ($P_1$) represents putative conditions not driven by current airline or ATM objectives but which may nevertheless benefit the passenger. This scenario, rebooking disrupted passengers at airports based on minimising delays at their final destination, produced very weak effects when current airline interlining hierarchies were preserved. When these restrictions were relaxed, under $P_2$, marked improvements in passenger arrival delay were observed, although at the expense of an increase in total delay costs per flight (due to passenger rebooking costs). Trade-off results have also been observed in a US model [8]: compared to the traditional ration-by-schedule rule, rationing by aircraft size (three priority queues: ‘heavy’, ‘large’ and ‘small’ aircraft) was shown to decrease the total passenger delay by 10%, with a 0.4% increase in total flight delay. Rationing by passengers on-board decreased total passenger delay by 22%, with only a 1.1% increase in total flight delay.

[4] Differences shown are statistically significant ($p < 0.05$; z-tests) and exceeded a minimum change threshold applied to avoid reporting artefactual results (typically set at approximately 2% of the baseline mean values; not applied to the ratio metrics).

| Core metric | Units | $N_1$ & $N_2$<br>Inbound prioritisation based on simple pax numbers, or on onward flights delayed | $P_1$<br>Passenger reaccommodated based on delay at final destination …<br>… preserving interlining hierarchies | $P_2$<br>… relaxing interlining hierarchies | $A_1$<br>Departures times based on cost minimisation (& consideration of ATFM delay) |
|---|---|---|---|---|---|
| Flight departure delay | mins / flight | no significant changes in current flight-centric metrics: stresses need for passenger-centric metrics | | | |
| Flight arrival delay | mins / flight | | | | |
| Departure delay of departure-delayed flights | mins / flight | | | | |
| Arrival delay of arrival-delayed flights | mins / flight | | | | |
| Pax departure delay | mins / pax | no significant changes under simple inbound scenarios driven by passenger numbers, or by numbers of delayed onward flights | revised passenger re-booking rules produce only weak improvements whilst current airline interlining rules are preserved, c.f. → | = | +0.4 |
| Pax arrival delay | mins / pax | | | -0.4 | -1.6 |
| Departure delay of departure-delayed pax | mins / pax | | | = | = |
| Arrival delay of arrival-delayed pax | mins / pax | | | -2.2 | -9.8 |
| Passenger value of time | Euros / pax | | | -0.2 | -0.7 |
| Non-passenger costs | Euros / flight | | | = | = |
| Per-flight pax hard cost | Euros / flight | | | +26 | -40 |
| Per-flight pax soft cost | Euros / flight | | | = | = |
| Total flight cost | Euros / flight | | | +26 | -39 |
| Total flight cost per minute of departure delay | Euros / min | | | = | -7.8 |
| Reactionary delay ratio | ratio | | | 49% | 51% |

Figure 2. Summary of core results.

Returning to the POEM model, the prioritisation process $A_1$, assigning departure times based on cost minimisation (with due consideration of ATFM delays), markedly improved a number of passenger delay metrics and airline costs, the latter determined by reductions in passenger hard costs to the airline.

One of the very few negative outcomes associated with $A_1$ was an increase of two percentage points in overall reactionary delay. This was manifested through relatively few flights and was introduced purposefully by airlines through the cost model (i.e. waiting for late passengers) such that the overall cost to the airlines decreased.

Under $A_2$ (results not shown) the addition of independent, cost-based arrival management (see Table I) apparently foiled the benefits of $A_1$ due to lack of coordination between departures and arrivals. This was also reflected through the finding that $A_2$ caused increased dispersion (standard deviations) of all the core metrics, and produced the highest reactionary delay ratio of 58%. Comparably, it has been shown [29] using US data, that arrival queuing delay at certain airports is associated with a net reduction of delay in the network as a whole, whilst queuing at others is associated with a net increase. Non-linear relationships were demonstrated. Arrival queuing may thus have a delay multiplier effect in the network.

The ratio of arrival-delayed passenger over arrival-delayed flight minutes (both pertaining to delays of greater than 15 minutes) was 1.5 for the $S_0$, $P_1$ and $P_2$ simulations for the baseline traffic day and the high delay day, rising to 1.9 for $S_0$ on the high cancellation day. Notably, $A_1$ for the baseline traffic day resulted in a minimum value of this ratio of 1.3. These values compare well with the range 1.6 – 1.7 cited in Section II.

The importance of using passenger-centric metrics in fully assessing system performance is clearly made through the results shown in Fig. 2, since the changes were not expressed through any of the currently-used flight-centric metrics at the common thresholds set. Scenario $A_1$ appears to hold particular promise and will be studied in particular, along with the corresponding baseline ($S_0$) results, in the next sections.

## B. *Delay propagation*

Reactionary delays and their causes are determined *a posteriori*. If several passengers were connecting from different flights and all of them were late, we only considered the most restrictive connection (in actual minutes) as the reason for the reactionary delay being induced. In this sense, one flight can delay many others, but any given flight can only be delayed by one previous flight (the most restrictive one). This graph is thus a (propagation) tree.

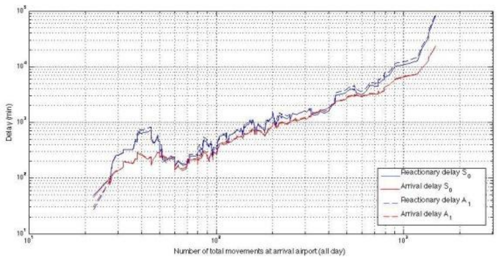


Figure 3. Arrival and reactionary delay, by airport size.

Fig. 3 shows total (daily) reactionary and arrival delay as a function of airport movements. Although large airports are associated with more reactionary and arrival delay, there is a considerable relative difference between these delay types at the smaller airports. For some of the forty smaller airports arrival delay was doubled (or even tripled) into reactionary delay. This is due to reduced delay recovery potential at such airports, for example through: flexible or expedited turnarounds; spare crew and aircraft resources (as yet not explicitly modelled in POEM); and, whether a given airport has sufficient connectivity and capacity to reaccommodate disrupted passengers. In practice, the business model of airlines operating at airports also influences these effects. Similar findings have been reported in some literature ([30], [31]).

Back-propagation (where an aircraft's outbound delay propagates back to an airport one or more times later in the day) was found to be an important characteristic of the persistence of delay propagation in the network. Paris Charles de Gaulle, Madrid Barajas, Frankfurt, London Heathrow, Zürich and Munich all demonstrated more than one hundred hours of back-propagated delay during the modelled (baseline) day. The prevalence of hub back-propagation has also been reported in the literature ([10], [31], [32]).

## C. *Granger causality directed network analysis*

Classical statistical instruments such as correlation analysis are only able to assess the presence of some common (equivalent) dynamics between two or more systems. However, correlation does not imply causality. Granger causality, on the other hand, is held to be one of the only tests able to detect the presence of causal relationships between different time series. It is an extremely powerful tool for assessing information exchange between different elements of a system, and understanding whether the dynamics of one of them is led by the other(s). It was originally developed by Nobel Prize winner Clive Granger [33] and although it was applied largely in the field of economics [34] it has received a lot of attention in the analysis of biomedical data ([35]-[37]).

A network reconstruction was computed for the flight and passenger layers for the $S_0$ and $A_1$ scenario simulations of the baseline traffic day, i.e. four reconstructions in total (the two baseline networks are shown in figures 4 and 5). The colour of each node represents its eigenvector centrality, from green (low centrality) to red (most central nodes). The size represents the out-degree, i.e. the number of airports that a given airport Granger 'forces' in terms of delay. The eigenvector centrality is a metric defined such that this centrality of a node is proportional to the centralities of those to which it is connected.

Comparing eigenvector centrality rankings through Spearman rank correlation coefficients showed [13] that all four network layers were remarkably different from each other ($r_s$: 0.01 – 0.07). These rankings demonstrated that different airports have different roles with regard to the type of delay propagated (i.e. flight or passenger delay) and that these were further changed under $A_1$. Indeed, a trade-off was introduced under $A_1$: the propagation of delay was contained within smaller airport communities, but these communities were more susceptible to such propagation. The absence of major hubs in the top five ranking lists for in-degree, out-degree and eigenvector centralities was evident. Indeed, the largest airports present in these rankings were Athens, Barcelona and Istanbul Atatürk. We previously reported similar findings in a network vulnerability analysis [2].

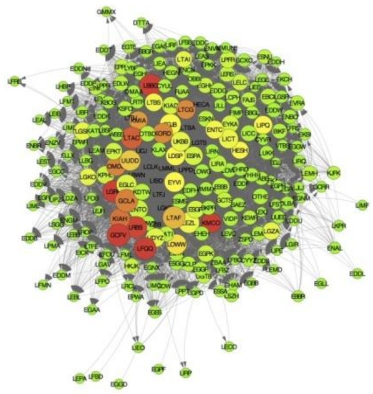

Figure 4. Flight delay causality network for $S_0$ simulation.

Investigating how congested airports form connected clusters in the US 2010 network, it was found [11] that the same airports were not consistently part of such clusters, implicating daily scheduling differences in delay propagation patterns. It was noted that being in the same cluster was a measure of correlation but not necessarily a sign of a cause and effect relationship. Notably, only two major hubs, Newark and

San Francisco, were present in the top ten for persistence in the largest congested clusters ([11]; "Supplementary information").

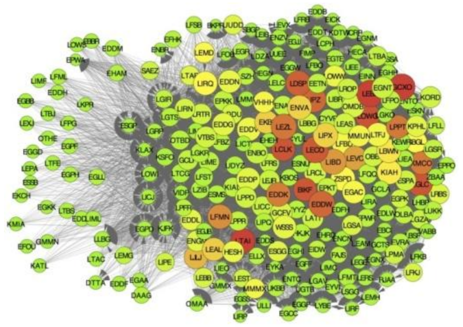

Figure 5. Passenger delay causality network for $S_0$ simulation.

## V. FUTURE RESEARCH

An examination of the socio-political, regulatory and technical contexts, and of the state of the art regarding current modelling, suggests that there is a role for the continued development of tools to explore the impacts of flight and passenger prioritisation strategies. The results we have presented, building the first explicit passenger connectivity simulation of the European air transport network, show that passenger-centric metrics, including appropriate network and cost considerations, are necessary complements to existing flight-centric metrics in order to fully evaluate system performance. These enable insights into such performance, in addition to oversight. Building on the POEM model's flexibility, we plan to implement higher fidelity en-route behaviour and ATFM modelling functionalities, and to use the tool to explore future market trends (such as traffic levels, aircraft size, load factors, service frequencies and hub wave structures), robustness under disruption, and the trade-offs between various prioritisation and policy strategies. These may be examined not only at the network level, for example in the context of SES and SESAR high-level targets, but also for given airline route clusters and airports.

## REFERENCES


[1] European Commission, Regulation (EC) No 261/2004 of the European Parliament and of the Council of 11 February 2004, Official Journal Vol. 47, February 2004.

[2] A. Cook, G. Tanner and M. Zanin, "Towards superior air transport performance metrics – imperatives and methods", Journal of Aerospace Operations, 2, 3–19, 2013.

[3] S. Bratu and C. Barnhart, "An analysis of passenger delays using flight operations and passenger booking data", Sloan Industry Studies Working Paper WP-2004-20, 2004.

[4] L. Sherry, D. Wang, N. Xu and M. Larson, "Statistical comparison of passenger trip delay and flight delay metrics", Transportation Research Board 87th Annual Meeting, Washington DC, 2008.

[5] D. Wang, "Methods for analysis of passenger trip performance in a complex networked transportation system", Doctoral thesis, George Mason University, Fairfax VA, 2007.

[6] G. Calderón-Meza, L. Sherry and G. Donohue, "Passenger trip delays in the U.S. airline transportation system in 2007", Third International Conference on Research in Air Transportation, Fairfax VA, 2008.

[7] L. Sherry, G. Calderon-Meza and G. Donohue, "Trends in airline passenger trip delays: exploring the design of the passenger air transportation service", Transportation Research Board 89th Annual Meeting, Washington DC, 2010.

[8] B. Manley and L. Sherry, "The impact of ground delay program (GDP) rationing rules on passenger and airline equity", Third international conference on research in air transportation, Fairfax VA, 2008.

[9] L. Sherry, "A model for estimating airline passenger trip reliability metrics from system-wide flight simulations", Journal of Transport Literature, 7(2), 319–337, 2013.

[10] N. Pyrgiotis, K. M. Malone, A. Odoni, "Modelling delay propagation within an airport network", Transportation Research Part C, 27, 60–75, 2013.

[11] P. Fleurquin, J. J. Ramasco and V. M. Eguiluz, "Systemic delay propagation in the US airport network", Scientific Reports 3, 1159, 2013.

[12] S. H. Kim, E. Feron, J-P. Clarke, A. Marzuoli and D. Delahaye, "Airport gate scheduling for passengers, aircraft, and operations", Tenth USA/Europe air traffic management research and development seminar, Chicago, 2013.

[13] SESAR, POEM (E.02.06) Deliverable 6.2 - Final Technical Report, September 2013.

[14] SESAR Consortium, SESAR Definition Phase: Milestone Deliverable 2, Air Transport Framework - The Performance Target, 2006.

[15] European Commission, Flightpath 2050 – Europe's Vision for Aviation (Report of the High Level Group on Aviation Research), ISBN 978-92-79-19724-6, DOI 10.2777/50266, 2011.

[16] European Commission, White Paper: Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system, Brussels, 2011.

[17] European Commission, Possible revision of Regulation (EC) 261/2004 on denied boarding, long delays and cancellations of flights, Roadmap Version 1, November 2011.

[18] European Commission, Air passenger rights revision (memo), Brussels, 13 March 2013.

[19] EUROCONTROL, Performance Review Report 2010: an assessment of air traffic management in Europe during the calendar year 2010, EUROCONTROL Performance Review Commission, 2011.

[20] EUROCONTROL, Performance Review Report 2012: an assessment of air traffic management in Europe during the calendar year 2012, EUROCONTROL Performance Review Commission, Brussels, 2013.

[21] Eurostat, Air passenger transport by reporting country, http://epp.eurostat.ec.europa.eu/portal/page/portal/transport/data/database, 'avia_paoc' dataset accessed September 2013.

[22] Performance Review Body of the Single European Sky, Proposed regulatory approach for a revision of the SES performance scheme addressing RP2 and beyond, EUROCONTROL Performance Review Commission, v.1.0, 2012.

[23] EUROCONTROL, PRB Annual monitoring report 2012 volume 1, European overview and PRB recommendations (Ed. 1), September 2013.

[24] International Civil Aviation Organization, Doc 9854, Global Air Traffic Management Operational Concept, First Edition, 2005.

[25] EUROCONTROL, PRB advice to the Commission in the setting of Union-wide performance targets for RP2 – Final Report, September 2013.

[26] SESAR, SESAR Concept of Operations Step 1 (Ed. 1), 2012.

[27] SESAR, SESAR Concept Of Operations Step 2 (Ed. 1), 2013.

[28] SESAR, European ATM Master Plan (Ed. 2), The Roadmap for Sustainable Air Traffic Management, 2012.

[29] I. Kwan and M. Hansen , "US flight delay in the 2000's: an econometric analysis", Transportation Research Board 90th Annual Meeting, Washington DC, 2011.

[30] S. AhmadBeygi, A. Cohn, Y. Guan and P. Belobaba , "Analysis of the potential for delay propagation in passenger airline networks", Journal of Air Transport Management, 14, 221– 236,2008.

[31] M. Jetzki, "The propagation of air transport delays in Europe", Doctoral thesis, Department of Airport and Air Transportation Research, RWTH Aachen University, Germany, 2009.

[32] N. Pyrgiotis, "A public policy model of delays in a large network of major airports", Transportation Research Board 90th annual meeting, Washington DC, 2011.

[33] C.W.J Granger, "Investigating causal relations by econometric models and cross-spectral methods", Econometrica, 37 (3), 424–438, 1969.

[34] K.D. Hoover, Causality in macroeconomics, Cambridge University Press, 2001.

[35] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura and S. L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality", Proceedings of the National Academy of Science USA, 101 (26), 9849–9854, 2004.

[36] M. Kamiński, M. Ding, W.A. Truccolo and S. L. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance", Biological Cybernetics 85 (2), 145–157, 2000.

[37] A. Roebroeck, E. Formisano and R. Goebel, "Mapping directed influence over the brain using Granger causality and fMRI", NeuroImage 25 (1), 230–242, 2005.