

Passenger-Oriented Enhanced Metrics

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—We report on a project building the first European ATM simulation combining flight and passenger trip data. New propagation-centric and passenger-centric performance metrics are described. The new metrics will be compared with existing, classical metrics, to compare their respective intelligibility, sensitivity and consistency. The trade-offs in performance across the metrics under a range of flight and passenger prioritisation scenarios will be examined. The corresponding regulatory and socio-political contexts are described. Complexity science techniques demonstrate the need to extend flight-centric network representations to include the passenger perspective.

Keywords—delay propagation; passenger-centric; metric.

Foreword—This paper describes a project that is part of SESAR Workpackage E, which is addressing long-term and innovative research. (The main model outputs of the Project are not yet available. We focus here on the metric design and some exploratory analyses.)

I. INTRODUCTION

The propagation of delay through the network remains a significant and costly operational challenge to air traffic management – yet we have virtually no metrics that specifically measure this. There is also a growing political emphasis in Europe on service delivery to the passenger, and passenger mobility, yet all our metrics are flight-centric rather than passenger-centric. How are we to measure the effectiveness of new passenger-driven performance initiatives in air transport in general, and ATM in particular, if we do not have the corresponding set of passenger-oriented metrics? The same question holds for efforts to understand and reduce delay propagation.

The ‘POEM’ (Passenger-Oriented Enhanced Metrics) project is under the SESAR Workpackage E theme: ‘mastering complex systems safely’. The goal of POEM is to build the first European ATM model, which combines flight and passenger trip data. Drawing in part on complexity science new performance metrics are also developed to explore delay propagation, which will vitally complement new passenger-centric metrics. The model then examines performance, measured through both new and existing metrics, under a range of flight and passenger prioritisation scenarios.

Our aim in this paper is to offer an integrated overview of the project to date, i.e. without focus on any particular aspect, since full details of each component are available in the dedicated deliverables.

In Section II we set the research, regulatory and socio-political contexts, and describe the approach. Section III presents the design landscape of the new metrics and proposes the model’s classical metrics. Section IV focuses on a selection of metrics used in network theory, then outlines some corresponding, initial analyses, using passenger flow and aircraft movement data. Due to limitations of space and the context of the paper within the Workpackage E theme dedicated to complexity science, we assume a basic knowledge thereof. Finally, a description of some of the key features of the model is presented in Section V. At the time of press, the model is nearing its first phase of implementation.

II. CONTEXT AND APPROACH

A. Overview of the state of the art¹

The average delay of a delayed flight and the average delay of a delayed passenger are not the same. The air transport industry is lacking passenger-centric metrics; its reporting is flight-centric. These metrics may even give contradictory results.

Dedicated metrics for propagation are also conspicuous by their absence. EUROCONTROL has pointed out [1] that there is ample scope for further research in this domain:

“Despite [...] the large share of almost 50% of reactionary delay, there is presently only a limited knowledge of how airline, airport and ATM management decisions affect the propagation of reactionary delay throughout the air transport network.”

“Reactionary delays are by definition a network issue and a better understanding of the contribution of airports, airlines and ANS towards those network effects and possible measures to mitigate those effects would be desirable, particularly with a view to the network manager that will be established under the SES II initiative.”

Using US historical flight segment data from 2000 to 2006, to build a passenger flow simulation model to predict passenger trip times, [2] 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 I).

¹ Space prohibits a full literature review here. An extensive review was presented in an earlier project deliverable (available on request from the authors) and variously updated in subsequent deliverables.

TABLE I. PREDICTED PAX TRIP DELAY BY PERFORMANCE CHANGES

<i>Performance change</i>	<i>Predicted pax trip delay change^a</i>
15-minute reduction in flight delay	-24% ^b
improved airline cooperation policy in re-booking disrupted passengers	-12%
flights cancelled earlier in the day	-10%
decreasing load factor to 70%	-8%

a. Source: [2].

b. With a concomitant saving of approximately USD 2.3 million in passenger value of time per day.

Using large data sets for passenger bookings and flight operations from a major US airline, it has been shown [3] how 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. These authors conclude that flight-leg delays severely underestimate passenger delays for hub-and-spoke airlines. Based on a model using 2005 US data, [4] concurs that “flight delay data is a poor proxy for measuring passenger trip delays”.

An “inherent flaw in the design of the passenger transportation service” has been pointed out [5], 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 [5]. Today, in neither the US nor Europe, are on-time performance and predictability sufficiently high to obviate the requirement for significant levels of passenger disruption recovery.

With regard to delay propagation and passenger-centric metrics in particular, ATM is faring better at gaining *oversight* than acquiring *insight*. This may also be said to apply to other key performance areas (KPAs). In order to improve service delivery further, we therefore first need to better characterise and measure performance, through improved metrics. The design of such metrics does not, of course, occur in a vacuum. The corresponding regulatory and socio-political contexts have to be taken into account. We will next briefly review these contexts up to 2012, before introducing the approach adopted in POEM, and moving on to present the metrics that we will explore in the model.

B. The Single European Sky performance scheme

In September 2010, EUROCONTROL accepted the European Commission’s designation as 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 II.

TABLE II. SES PERFORMANCE SCHEME REFERENCE PERIODS

<i>Reference period</i>	<i>Applicable years</i>
RP1	2012 - 2014
RP2	2015 - 2019
RP3	2020 - 2024

RP1 addresses mainly the en-route part of air navigation service provision, focusing target-setting on en-route capacity, environment and cost efficiency. EU-wide performance targets for RP1 were adopted by the European Commission in February 2011 and work on RP2 preparation was launched by the PRB in June 2011. Proposals to improve and reinforce the performance scheme from RP2 onwards have been set out by the PRB. The proposals aim to support greater consistency between the performance scheme and other SES functionalities, such as the charging scheme, functional airspace blocks (FABs) and the deployment of SESAR technology, as well as with other EU policies. RP2 also 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 [6]. Including full consultation processes, the aim is to finalise the amendment of the performance scheme and the charging scheme by the end of 2012, to allow EU-wide performance targets for RP2 to be agreed for these KPAs before the end of 2013.

Current ATM key performance indicators (KPIs) in Europe are (inevitably) rather high-level. Many targets have been set [7] at the European level. For capacity (under RP1), air traffic flow management (ATFM) en-route delay per flight (with a weather delay allowance managed at the network level) has a target of 0.5 minutes by 2014 (with incentives set on the Network Manager and FABs). 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). The fifteen minute threshold for defining arrival and departure delay has, historically, been common to both Europe and the US. SESAR’s Performance Target [8] refined these significantly, however, as shown in Table III.

TABLE III. SESAR PERFORMANCE OBJECTIVES AND TARGETS

<i>SESAR metric</i>	<i>Target for 2020</i>
departure punctuality	≥ 98% of flights departing as planned ±3 mins other 2%: average delay ≤ 10 mins
arrival punctuality	> 95% of flights arrival delay ≤ 3 mins 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 $\sigma < 1.5\%$ of route mean ^a

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

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. It would be easier to reach the reactionary delay target of a 50% reduction by 2020, relative to 2010, with appropriate metrics to enlighten us regarding propagation mechanisms and ‘hot-spots’ (delay multiplier nodes in the network). We will propose such metrics, later in this paper, since the industry is currently missing metrics that offer any real insight into these mechanisms.

C. Recent ATM performance

It is somewhat difficult to accurately judge *underlying* progress towards these targets. In 2011, European air traffic remained below the pre-economic crisis levels of 2008: the year that marked the end of a sustained period of growth. However, one positive, and clearly related, aspect of lower levels of traffic in 2011, was that arrival punctuality improved. 18% of arrivals were delayed by more than 15 minutes, compared with just over 24% in 2010 [9]. (However, 2010 was itself a poor year, largely due to ATC strikes and extreme weather.) ANS contributed through a significant reduction in total ATFM delays, mainly driven by a reduction of en-route delays.

The ratio of reactionary delay to primary delay in Europe grew steadily from 2003 to 2008. Since then, it has roughly levelled off. Reactionary delay represented 45.8% of all delay in 2011, compared to 46.7% in 2010, although this accompanied a fall of one third in reactionary minutes per flight, due to the overall fall in delay minutes in 2011 [9].

The real test will be to see how such metrics perform when Europe emerges from its current economic situation, and traffic picks up significantly. With financial pressures mounting in difficult times, often driven by increasing shareholder scrutiny in a privatised context, and as the reach of the SES performance scheme is strengthened through legislative measures, focus on all of these metrics will increase.

Although 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, this nevertheless remains pertinent in terms of service delivery to the passenger, to which we turn in the next section.

D. Socio-political context

SESAR’s ‘Performance Target’ [8] refers frequently to the concept of society and the passenger. The ‘societal outcome’ cluster of KPAs², 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’ cluster³ 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’ [10] and the European Commission’s 2011 White Paper (‘Roadmap to a Single European Transport Area’, [11]). Metric design also needs to reflect the progress of corresponding planned regulatory review, particularly with regard to the underpinning regulatory instrument, Regulation 261 – the European Union’s air passenger compensation and assistance scheme [12].

A roadmap [13] for the possible revision to Regulation 261 was published in late 2011. A specific example of the need for metrics to take account of changing regulation is the potential extension [13] of the legislation to cover passengers’ missed connections, which is neither covered by current law nor current metrics.

A consultation on the potential revision was completed in March 2012. There was little consensus on the way forward, with responses from airlines and consumer/passenger organisations often directly opposed [14]. In May 2012, stakeholders met at a conference on air passenger rights to discuss the consultation findings.

The Commission intends to put forward a proposal to revise the Regulation by the end of 2012, making the current Regulation more effective, without imposing undue burdens on operators or passengers [15].

E. Project structure and rationale

The POEM project is supported by a consultation and dissemination workpackage. This included an on-line user-requirements survey addressing KPAs. This stakeholder survey secured 157 responses from airlines, airport authorities, air navigation service providers (ANSPs), civil aviation authorities, EUROCONTROL, Regulation 261 national enforcement bodies and researchers/academics.

It was followed by a complementary one day seminar and workshop in central London in January 2012, attracting approximately 60 delegates, plus liaison with the PRB with regard to the ongoing Performance Scheme consultation.

At the core of POEM is the design of new metrics and the evaluation of these through a European network simulation model under different flight and passenger prioritisation scenarios. The prioritisation scenarios provide primary inputs into the network simulation model and were designed in parallel with the metrics presented in Section III. This design was informed by both the literature review and the stakeholder consultation process.

Fig. 1 shows the model architecture of POEM, with a focus on the scenarios and metrics and the relationships between them. The central oval represents the ATM system: the main model. The prioritisation scenarios applied in the model may be classified by three central themes according to the agency/orientation of the decision-making. Each is figuratively represented as a ‘stream’ (horizontal grey band) with its impact ‘flowing into’ the ATM system.

² Environment, safety, security.

³ Capacity, cost effectiveness, efficiency, predictability, flexibility.

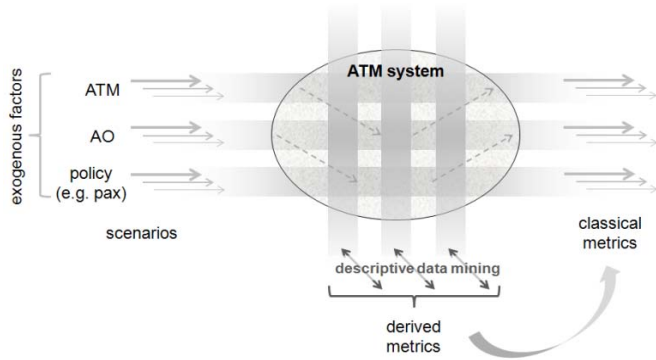


Figure 1. Project architecture: scenarios and metrics.

ATM (ANSP) and AO scenarios involve decision-making based on reasonable information for that agent to possess in either the current information environment, or a future one (e.g. in the context of System-Wide Information Management). A policy-driven scenario represents the special case where we run the model under putative conditions not driven by current airline or ATM objectives, but which may nevertheless benefit the passenger. In addition to this, over-arching all the scenario streams, are exogenous factors, such as economic growth or recession, which may drive traffic volumes up or down, and which are largely independent of airline or ATM practice, and most air transport policies. These are out of scope for POEM, although they are partly reflected in the different passenger and traffic levels of the four months to be modelled (see Section V).

The multiple arrows in each scenario stream represent the different levels of each scenario, as described in Section V(B). ‘Level 0’ represents the current, common baseline situation and increasing level numbers represent increasing levels of intervention with respect to the current baseline.

For each of the scenario streams, we may expect a corresponding output effect (on the right-hand side of the figure) on the metrics described in sections III and IV. For example, if a scenario in the ATM (ANSP) stream prioritises flights according to aircraft delay minutes, we would naturally observe a reduction in metrics such as average flight delay. If a scenario in the AO stream prioritises flights according to passenger cost of delay to the airline, a reduction in passenger metrics associated with delay cost will be observed. More interesting, however, are the quantitative effects observed *between* these streams (represented by the dashed arrows between the horizontal bands). It would be expected that prioritising flights according to aircraft delay minutes (as a scenario in the upper stream) will also reduce passenger delay costs (a metric in the second stream) – in fact, this relationship would be expected to be superlinear, since the relationship between aircraft delay minutes and the corresponding cost of passenger delay to the airline is also superlinear.

Also to be investigated is how the different levels of intervention applied in the scenarios affect the ‘levels’ of the metrics. ‘Increasing levels’, used to describe the metrics, may

be used to identify higher functions, such as variance and kurtosis, as opposed to simple averages (see also Section III(C)).

III. NEW METRICS FOR ATM

A. Expanding the metric landscape

We define a metric as any quantitative measure, particularly one which usefully expresses some output of a system (usually performance), part of the system, or (an) agent(s) within it, usually over an aggregate scale and often as a ratio (e.g. per flight). Fig. 2 shows a metric classification, which we define in order to more clearly present the manner in which their scope needs to be extended in ATM and to differentiate between the types. The term ‘classical’ metrics is used to denote those that are pre-defined (such as average aircraft delay), are univariate (draw on one variable in the data) and do not use complexity science techniques. Some of these types of metric are already commonly in use (such as, indeed, average aircraft delay) whilst others are not (such as average passenger delay) – and, arguably, thus conspicuous by their absence.

‘Non-classical’ metrics defines both (non-complexity) ‘derived’ metrics, which are in contrast to the classical metrics in that they are not (fully) pre-defined but are derived from the data iteratively and are typically multivariate, and those drawn from complexity science. An example of a derived metric is a factor obtained as the result of factor analysis. (We will use this method, and variants thereof, such as principal component analysis, which is especially good at dealing with the issue of multi-collinearity, but we do not discuss this in this paper).

An example of a simple complexity metric is the degree of a node – we will discuss this, and others, in Section IV. Fig. 2 shows that these relationships are not wholly mutually exclusive. Data mining techniques may be applied not only to generate non-classical metrics but also in topology characterisation, such as identifying complex network communities (groups of densely connected nodes sharing only few connections with nodes outside their group). These techniques are not needed to define classical metrics, however.

Whilst it is thus relatively straightforward to identify some metrics that belong decisively to one of the categories, the *overlap* between the categories is less well defined and is of particular interest to explore. For example, how well do non-complexity metrics and methods capture certain features of ATM system dynamics (such as delay propagation) compared with those of complexity science?

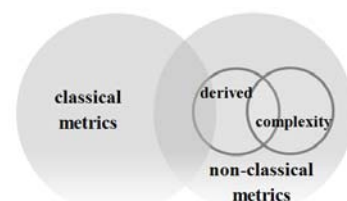


Figure 2. Relationships between metric types.

Such answers may help to compellingly stress the specific benefits of complexity techniques, by throwing the outcomes into focus with non-complexity methods, thus allowing researchers in ATM to propose more specific benefits for other disciplines and to foster improved outreach beyond the field. Such a meta-methodological approach also mitigates what is sometimes referred to as ‘research enculturation’, whereby a field of research adheres too narrowly to its own received wisdoms and culture.

Returning to the SES context, we note that specifically in this context [9], a performance indicator (PI) refers to an indicator used for the purpose of “performance monitoring, benchmarking and reviewing”, whereas a key performance indicator (KPI) is for performance target setting. KPIs need to be chosen that are intelligible (preferably to the point of being simple), sensitive (in that they accurately reflect the aspect of performance being measured) and consistent (we cannot refine them from one period to another without losing comparability). The concomitant disadvantages are that it is difficult to adapt them in response to new data or methods, and that they may not afford the best understanding of system dynamics. Trade-offs between these desirable properties often have to be made.

B. Benefits of non-classical metrics

It is explained in [16] how much of modern science is based on first-principle models to describe systems, starting with a basic model (such as Maxwell’s equations for electromagnetism, only later empirically proven), which are then verified (or otherwise) by experimental data to estimate some of the parameters. However, in many domains, such first principles are not known, and/or the system is too complex to be formalised mathematically. Through data mining, there is currently a “paradigm shift from classical modelling and analyses based on first principles to developing models and the corresponding analyses directly from data.”

Of the non-classical metrics, identified in Fig. 2, it is the complexity metrics on which we will focus in this paper. Data mining has much to offer the field of complexity science, not least in the development of performance metrics. Due to the very nature of derived metrics, we are more likely to make unexpected findings, and to deepen our understanding through being prompted to explain counterintuitive results, than we would through the use of classical metrics alone and without the context of complexity science. Some important analogies emerge between the choice of factors in a factor analytic model and the choice of nodes in a graph theoretic model, for example.

The double-headed arrows in Fig. 1, representing the descriptive data mining methods employed in POEM, are double-headed to denote their iterative nature. The collective grouping across the vertical bands denotes the multivariate methods of the solutions and the fact that these are not necessarily aligned with the main streams of the classical metrics.

We stress the complementary approach adopted in POEM, across metric types, and turn next to the classical context.

C. Holistic approach – value of classical metrics

The following two tables present the classical metrics to be used in the model, based on dedicated literature reviews and internal design (full reporting available from the authors). Table IV shows the classical, propagation-oriented metrics. Classes shown in *italics* indicate that the metrics are primarily driven by delay, as compared with propagation *per se*, although the two phenomena are obviously closely related.

TABLE IV. PROPAGATION-ORIENTED METRICS

<i>Metric class</i>	<i>Type</i>
delayed departures	<i>count</i>
delayed arrivals	<i>count</i>
departure delay	<i>duration(s)^a</i>
arrival delay	<i>duration(s)^a</i>
cancelled flights	<i>count</i>
extra flight time	<i>duration(s)</i>
extra gate time	<i>duration(s)</i>
reactionary minutes	<i>duration(s)</i>
back-propagation ^b	<i>ratio</i>
reactionary disruptions	<i>count (of disrupted flights^c)</i>
reactionary depth / disruptions	<i>ratio of nodes/flights</i>
reactionary / primary delay ratio	<i>ratio of durations</i>
reactionary depth	<i>count (of disrupted flights^c in the longest path in propagation tree)</i>

a. To specifically include % above certain thresholds.

b. Ratio of reactionary delay from an airport that later propagates back to the same airport.

c. Excluding the causal flight.

Table V shows the classical, passenger-oriented metrics. Classes shown in *italics* indicate that the metrics are also directly linked to value of time evaluations (see Section V(A)). Each such metric is evaluated on a passenger per-trip basis. Specific metrics will be used within each class, such as departure delay average and variance.

TABLE V. PASSENGER-ORIENTED METRICS

<i>Metric class</i>	<i>Type</i>
delayed departures	<i>count</i>
delayed arrivals	<i>count</i>
departure delay ^b	<i>duration(s)^a</i>
arrival delay	<i>duration(s)^a</i>
final arr. delay / scheduled trip time	<i>ratio</i>
cancelled flights	<i>count</i>
missed connections	<i>count</i>
re-routes	<i>count</i>
extra flights	<i>count</i>
extra flight time	<i>duration(s)</i>
weighted load factor ^c	<i>value</i>
aborted trips	<i>count</i>
extra wait time	<i>duration(s)</i>

a. To specifically include % above certain thresholds.

b. Estimates excess wait time at airports.

c. Weighted by flight durations: very crude estimate of pax comfort.

Different scales of measurement, and levels of disaggregation (e.g. by airline and airport type), will also be applicable within each class. Many of the metrics will be determined in terms of their associated costs (to the airlines). However, whilst the cost impacts of different scenarios will be examined, we cannot evaluate a cost *efficiency* KPI for them, since the costs of implementing the prioritisation scenarios are not assessed.

POEM will embrace the *distribution* of metrics rather than focusing, as typically practiced, on point estimates and central tendencies (which tell us nothing about predictability – one of the eleven ICAO KPAs adopted by SESAR [8]). Considering passenger delay cost to the airline, for example, the mean cost (μ_c), the standard deviation of the cost (σ_c) and excess kurtosis (k'_c) of the cost, will all be considered, and may well be sensitive to the different scenarios in different ways. Such measures will contribute, in part, to the ‘levels’ of the (classical) metrics (mentioned in Section II(E)).

Returning to the discussion of Section III, simple averages may be (apparently) intelligible and consistent, but not sensitive: a deterioration in one part of the system may partially offset an improvement elsewhere, resulting in only a small change in the mean value. Further evidence for the importance of considering dispersion arises from the fact that heavily skewed distributions of passenger trip delay demonstrate that a small proportion of passengers experience heavy delays, which is not apparent from flight-based performance metrics ([2], [17]). Also of note, the dispersion metric ‘% of flights delayed by more than a certain amount of time’ scored particularly highly in terms of usefulness, in the on-line stakeholder survey mentioned earlier.

There will not only be collinearities within certain metric classes, but also between some of them (e.g. departure delay is the main driver of arrival delay in Europe ([1], [9])), including between the propagation- and passenger-oriented metrics. In other cases, interesting *negative* correlations may be observed: for example, comparing different rationing rules in a model ground delay programme rationing rule simulator, it was found [18] that passenger delays could be significantly decreased with a slight increase in total flight delay.

IV. COMPLEX METRICS, NETWORK REPRESENTATIONS

This section first presents a selection of metrics used in network theory, then outlines some differential, exploratory analyses using passenger flow and aircraft movement data.

A. Complexity metrics – definitions

1) Degree

The number of connections a node has, or, in other words, the number of neighbours; the greater the degree, the more important that node is, functionally, within the network. When nodes are defined to represent some parameterisation of delay, for example, if we had a few nodes with a very high degree, this would suggest that those nodes were responsible for the propagation of delay in the network. On the other hand, if all

nodes had more or less the same degree, no delay multiplier node would be suggested.

2) Betweenness

The number of shortest paths (taking into account all pairs of nodes) which pass through a node; nodes with high betweenness are usually those nodes that connect different communities, e.g. in the ATM context allowing perturbations to spread between different parts of the system.

3) Link density [19]

The number of links in the network, l , divided by the maximum number of links that could be present; for a network composed of n nodes, the link density is thus $l/(n(n-1))$.

4) Degree-degree correlation [19]

Pearson’s correlation coefficient between the degrees of pairs of nodes connected by a link; correlations > 0 indicate the presence of assortativity (e.g. hub-hub connections).

5) Global efficiency

The ease of information flow between pairs of nodes; the (generic) cost of this communication can be approximated by the distance (length) of the shortest path connecting two nodes – the normalised global efficiency is defined as the mean value of the inverse of such distances, d , as given by (1).

$$E_g = 1/\{n(n-1)\} \sum_{i,j \neq i} 1/d_{i,j} \quad (1)$$

B. Analysis of three network representations

For global efficiency ([20], [21]), the distances may be defined as required (e.g. Great Circle distances could be used for flights), although they are often treated as a topological distance, i.e. based on the number of links needed to ‘move’ from one node to another. We adopt such a topological treatment here, which will incorporate shortest paths.

We have evaluated the three *network-level* metrics, of Section IV(A), for three distinct network representations. In each case, European airports were the nodes, but the links were defined differently. The same data were used as described in Section V(A), for September 2010, although only in a relatively crude manner, at this stage.

Firstly, for the air traffic movement data, each link has a weight proportional to the number of direct flights operating between the two nodes (‘flights’, Table VI). This network represents the physical layer upon which passenger transportation takes place. Secondly, a passenger origin-destination network was created from the passenger data (itineraries were truncated within Europe). Each link has a weight proportional to the number of passengers travelling between each pair of airports, but independent of the actual routing (‘passenger O-Ds’, Table VI). Thirdly, a disaggregate passenger network was created based on all passenger legs flown, each link proportional to the number of passengers flying that leg (‘passenger legs’, Table VI).

TABLE VI. NETWORK METRICS FOR THREE REPRESENTATIONS

<i>Network</i>	<i>Link density</i>	<i>Degree-degree correlation</i>	<i>Global efficiency</i>
Flights	0.03	0.05	0.81
Passenger O-Ds	0.12	-0.06	0.94
Passenger legs	0.03	0.03	0.93

Values quoted to 2 d.p.

In Table VI, at the aggregate level of metric, we note that the networks have similar characterisations. As expected, the link densities are very similar for the flights and passenger legs (there is a small difference beyond the second decimal place due to data coverage) and rather higher for the passenger O-Ds. The degree-degree correlations are approximately zero (loosely interpreted as hubs connecting to spokes as frequently as to other hubs) and the global efficiencies are higher for the two passenger layers, where load factors and aircraft sizes implicitly contribute to the more naïve flight-based representation.

In a vulnerability analysis for each of the three network representations, each node (airport) was removed, in turn, and the global efficiency of the whole network re-calculated without that node (Table VII). The node's flights and (first order) passengers were removed, as if from a sudden closure. The simple 'flights' representation naïvely considers node-node connections in the complete absence of passenger data. This is indeed the basis of much current research in this area ([22] – [26]) where neither of the important passenger layers is considered. The 'flights' layer only captures point-to-point passenger movements, and even then only partially, because neither aircraft size nor load factors are considered. The airports in the first column are all hubs (except Toussus-le-Noble). The presence of these airports in the list may be interpreted in terms of the high number of *direct* destinations they serve, which are relatively poorly covered by near-neighbours. The apposite observation is the difference between the three lists, and the absence of 'obvious' candidates such as Heathrow and Schiphol from the particular perspective of network vulnerability.

TABLE VII. TOP TEN CRITICAL AIRPORTS BY NETWORK TYPE

<i>Flights</i>		<i>Passenger O-Ds</i>		<i>Passenger legs</i>	
Athens	4.12	Stockholm	3.79	Stockholm	6.04
Istanbul Atatürk	3.21	London Stansted	3.33	Oslo	4.60
Madrid	2.33	Oslo	3.04	Paris CDG	4.37
Paris CDG	2.17	London Gatwick	2.04	Helsinki	4.20
Paris Orly	2.07	Copenhagen	1.70	Istanbul Atatürk	3.16
Rome Fiumicino	1.95	Toulouse	1.59	Athens	2.59
Lisbon	1.54	Helsinki	1.55	London Stansted	2.49
Toussus-le-Noble ^a	1.39	Palma de Mallorca	1.32	Frankfurt	2.42
Prague	1.24	Madrid	1.26	Madrid	2.26
Vienna	1.18	Tromsø	1.25	Paris Orly	2.14

Percentage falls in global efficiency shown, to 2 d.p.

a. General aviation airport, with flights to/from 140 destinations (September 2010).

The importance of modelling the network from a passenger-centric perspective, in terms of both investigating network properties and designing metrics, and the shortcomings of treating the European network as a system of independent flights, are apparent. (A fuller methodology and interpretation of Table VII are available from the authors in a parallel paper recently completed.)

V. NEXT STEPS – IMPLEMENTING THE MODEL

A. Key model features

The POEM model will, for the first time, integrate passenger connectivity data into a full European ATM simulation. We will use EUROCONTROL traffic – PRISME – data and International Air Transport Association (IATA) – 'PaxIS' – passenger data. The data management process was started first under a dedicated workpackage due to the size of the task. This has also enabled us to better manage various data omissions (such as robust taxi-in times) and process modelling requirements (such as aircraft turnaround times). Both the allocation of passengers to aircraft (a significant advance compared with the preliminary analyses described in Section IV) and the implementation of the scenarios in the model, have necessitated the formulation and codification of a number of detailed and interconnected rules. These include realistic simulations for missed connections (such as dynamic passenger reaccommodation onto aircraft with free seats, using detailed fleet and load-factor data) and tail-tracked, aircraft wait rules. The model will cover four busy months, free of exceptional incidents (August and September 2009; August and September 2010) for 200 European and 50 extra-European airports. The model is a time-line graph – a random network with some stochastic elements built into most of the rules. First results are expected at the end of November 2012. Two airline case studies have focused on developing and testing specific aspects of the model rules, examined in an operational context. This included a dedicated workshop at London Gatwick. The model will be calibrated using independent data sources, i.e. additional data not used in the derivation of the model. These will include, for example: passenger and traffic volumes for most of the airports; high-level data on delay distributions and cancellation rates; plus, operational data from the case studies.

A significant advance on earlier work will be the explicit estimation of reactionary costs (since each flight is individually modelled with its connectivity dependencies) and of the passenger costs of delay to the airline (for example, based on Regulation 261, interlining hierarchies, ticket types and IATA proration rules). In previous work, all of these costs were estimated statistically. Passenger value of time will be quantified as a function of delay at the final destination only; insufficient data were available to assign wait and travel time penalties.

B. Model scenarios

Table VIII summarises the scenarios introduced in Section II(E). Note that the ATM scenarios are ANSP-moderated and the policy-driven scenarios are bolder than the current scope of Regulation 261.

TABLE VIII. SCENARIO TYPES, LEVELS AND OUTLINE DESCRIPTIONS

Type, level	Outline description
ANSP, 1	Prioritisation of inbound flights, based on simple passenger numbers
ANSP, 2	Prioritisation of inbound flights, based on no. of onward flights delayed by connecting inbound pax
AO, 1	Departure slots allocated based on delay costs – if ATFM delays are not severe, implement wait rules for premium passengers, long-haul passengers and minimum passenger load
AO, 2	Departure slots <i>and</i> arrival sequences based on delay costs – scenario ‘AO,1’ is implemented <i>and</i> flights are (independently) arrival-managed based on delay cost
Policy, 1	Passengers are reaccommodated based on prioritisation by arrival delay, instead of by ticket type, but preserving interlining hierarchies
Policy, 2	Passengers are reaccommodated based on prioritisation by arrival delay, regardless of ticket type, and also relaxing all interlining hierarchies

Discussing the correspondence between the scenarios and metrics, earlier, we observed how the relationships *between* the streams of Fig. 1 are anticipated to be the more interesting phenomena.

Furthermore, under one scenario, a given metric, M_1 , may be the most sensitive to a series of delay phase transitions in the network, whilst under another scenario, M_2 may be the better metric. In addition to looking for such robust metrics across ranges of scenarios, it may also be possible to derive further metrics, possibly (factorial) combinations of others, which are, by design, more robust across scenarios. By this, we mean that they display criterion validity (i.e. the metric’s value correlates with another key criterion (dependent) variable). An associated issue has arisen in some case study analyses (not shown; to be published), whereby we developed a new metric for passenger delay propagation, based on actual airport flight and passenger connectivity data. The new metric was sensitive to the frequency of aircraft reported late due to delayed boarding (a variable not in the original model), unlike a typical flight-centric metric, which was insensitive to this frequency. This problem was driven by the coding of early flights as zero-delay, an issue which also needs careful treatment in factor analyses.

The considered, specific derivation of robust metrics such as these will, it is hoped, be particularly useful in informing the general design of new metrics in ATM, thus promoting better capture of a wider range of system performance attributes, and overcoming the problems associated with some existing metrics.

It is hoped that the new propagation-centric and passenger-centric metrics will offer both contrasting and complementary insights into ATM performance. The new metrics will also be compared with existing, classical metrics, to compare their respective intelligibility, sensitivity and consistency. Of particular interest will be the anticipated trade-offs in performance across the range of metrics under the various passenger and flight prioritisation scenarios. It is hoped that such results will take ATM one step nearer to improved performance foresight as new technologies and solutions are introduced in SESAR.

REFERENCES

- [1] EUROCONTROL, Performance Review Report 2010: an assessment of air traffic management in Europe during the calendar year 2010, EUROCONTROL Performance Review Commission, 2011.
- [2] D. Wang, “Methods for analysis of passenger trip performance in a complex networked transportation system”, Doctoral thesis, George Mason University, Fairfax VA, 2007.
- [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] 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.
- [6] International Civil Aviation Organization, Doc 9854, Global Air Traffic Management Operational Concept, First Edition, 2005.
- [7] 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.
- [8] SESAR Consortium, SESAR Definition Phase: Milestone Deliverable 2, Air Transport Framework - The Performance Target, 2006.
- [9] EUROCONTROL, Performance Review Report 2011: an assessment of air traffic management in Europe during the calendar year 2011, EUROCONTROL Performance Review Commission, 2012.
- [10] 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.
- [11] European Commission, White Paper: Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system, Brussels, 2011.
- [12] 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.
- [13] European Commission, Possible revision of Regulation (EC) 261/2004 on denied boarding, long delays and cancellations of flights, Roadmap Version 1, November 2011.
- [14] European Commission, Public consultation on the possible revision of Regulation 261/2004 – results, report prepared by Steer Davies Gleave, 2012.
- [15] European Commission, Minutes for the stakeholder conference on the possible revision of Regulation 261/2004, Brussels, May 2012.
- [16] M. Kantardzic, Data mining – concepts, models, methods, and algorithms, (2nd Ed.), John Wiley & Sons, 2011.
- [17] 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.
- [18] 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.
- [19] L. da F. Costa, F.A. Rodrigues, G. Travieso and P.R. Villas Boas, “Characterization of complex networks: a survey of measurements”, Advances in Physics, 56 (1), 167–242, 2007.
- [20] V. Latora and M. Marchiori, “Efficient behavior of small-world networks”, Physical Review Letters, 87 (19), 198701, 2001.
- [21] P. Crucitti, V. Latora, M. Marchiori and A. Rapisarda, “Efficiency of scale-free networks: error and attack tolerance”, Physica A, 320, 622–642, 2003.
- [22] R. Guimerà, S. Mossa, A. Turtleschi and L.A.N. Amaral, “The worldwide air transportation network: anomalous centrality, community structure, and cities’ global roles”, Proceedings of the National Academy of Sciences of the USA, 102 (22), 7794–7799, 2005.
- [23] Z. Xu and R. Harriss, “Exploring the structure of the U.S. intercity passenger air transportation network: a weighted complex network approach”, GeoJournal, 73 (2), 87–102, 2008.
- [24] J. Wang, H. Mo, F. Wang and F. Jin, “Exploring the network structure and nodal centrality of China’s air transport network: a complex network approach”, Journal of Transport Geography, 19 (4), 712–721, 2011.
- [25] M. Zanin, L. Lacasa and M. Cea, “Dynamics in scheduled networks”, Chaos, 19 (2), 023111, 2009.
- [26] L. Lacasa, M. Cea and M. Zanin, “Jamming transition in air transportation networks”, Physica A, 388 (18), 3948–3954, 2009.