Modelling the airline costs of delay propagation

Andrew Cook and Graham Tanner
Department of Transport Studies
University of Westminster
London, United Kingdom
airspace-research@westminster.ac.uk

Abstract— Reactionary delays constitute nearly half of all delay minutes in Europe. A capped, multi-component model is presented for estimating reactionary delay costs, as a non-linear function of primary delay duration. Maximum Take-Off Weights, historically established as a charging mechanism, may be used to model delay costs. Current industry reporting on delay is flight-centric. Passenger-centric metrics are needed to better understand delay propagation. In ATM, it is important to take account of contrasting flight- and passenger-centric effects, caused by cancellations, for example. Costs to airlines and passenger disutility will both continue to be driven by delay relative to the original schedule.

Keywords- cost of delay, delay propagation, reactionary delay, Maximum Take-Off Weight, passenger-centric metric

I. INTRODUCTION

A. The cost of delay propagation

Primary delays associated with one aircraft often cause knock-on effects in the rest of the network. These are known as ‘reactionary’ (or ‘secondary’) delays. It is becoming increasingly important in ATM to understand delay costs, in order to better manage them. This paper presents a model for such costs, including a new model for the contribution from reactionary delays, which have shown a general upward trend in Europe since 2003 (EUROCONTROL, 2011). More than 24% of arrivals were over 15 minutes late in 2010, which was the worst on record (i.e. since 2001, ibid.).

Both the Single European Sky ATM Research (SESAR) and Next Generation Air Transportation System (NextGen) programmes identify opportunities to advance the state of the art with regard to flight planning and trajectory management. At the core of SESAR’s new operational concept is the ‘4D trajectory’ (SESAR Consortium, 2007). Similarly, NextGen describes (Joint Planning and Development Office, 2010) one of its eight key concepts as ‘trajectory-based operations (TBO)’. Within SESAR, the 4D trajectory is also referred to as the ‘business trajectory’. The airline will ‘own’ the business trajectory and have primary responsibility for its operation. When constraints do not allow the optimal choice for the airline, finding an alternative business trajectory, which achieves the “best business outcome” (SESAR Consortium, 2009) is left to the individual user and agreed through collaborative decision making to best balance demand with capacity. The precise definition of “best business outcome” is yet to be defined through research.

With regard to NextGen, under the grouping of ‘Trajectory and Performance-Based Operations and Support Policy Issues’, the JPDO provides (Joint Planning and Development Office, 2011) information on Policy Issue 0077, viz. ‘High Density Operations - Flight Prioritization’, citing a policy completion date of 2014: “Policies should be developed to set a construct or regime for prioritizing flights in congested operating environments. […] Will prioritization of operations be based on factors other than aircraft operating characteristics, such as aircraft capacity, operator mission? Will market or other ranking mechanism apply?”

If such business considerations are to address the cost of delay, these should include reactionary delay costs. We will first outline previous US and European research, before presenting our cost model.

B. Previous research in the US

Ferguson et al. (2011) developed a US model, based on the report by Cook et al. (2004), extending this to over 100 aircraft types by modelling (non-passerger) coefficients for the cost elements using US airline cost data. In general, good fits are reported and similar results are produced to the 2004 report, although one major difference between the two sets of results is that in the US model “the [passenger] delay cost coefficient is set to 0, since in the US, it is not incurred by the airlines”.

Another major consolidated source of airline delay costs in the US is a study sponsored by the Federal Aviation Administration (FAA) to estimate the total economic impact of flight delay (Ball et al., 2010). The costs were estimated by modelling the relationship between airline total cost (as opposed to flight-by-flight) and operational performance metrics. Increases in operating costs to airlines due to tactical delay (“delay against schedule”) and strategic delay (as “schedule padding”) are calculated using statistical cost models with airline data. The costs of schedule buffer were estimated using less impeded block times. In contrast to the model of Ferguson et al. (2011), costs to passengers are also calculated, as are wider costs such as macro-economic impacts.

In some recent examples of US cost estimations, USD 53 per minute for the cost of delay has been used (Klein et al., 2009) based on the direct operating costs of carriers. A direct operating cost of USD 65.80 per block minute has been used (Shepley, 2009), based on Air Transport Association of America (ATA) data. In a generic cost function (Pourtraklo and Ball, 2009) based on direct operating costs and passenger value...
of time (both sourced from ATA), the cost of delayed passengers to airlines was considered linear and approximated as one sixth of passengers’ value of time, capped after 15 hours; the first 15 minutes of delay of the cost function (airline and passenger costs) had a zero cost (similar to Cook et al., 2004). Cost parameters have also been developed (Zhang and Hansen, 2009) that include the passenger value of time at USD 28.60 per hour and an aggregate flight delay cost of USD 2,000 per hour (based on Cook et al., 2004).

C. Previous research in Europe

A consolidated estimation of European delay costs by aircraft type and phase of flight was first published in 2004 (Cook et al., 2004). It also contained a review of the preceding European literature. This work was updated in 2011 (Cook and Tanner, 2011). The values published in 2004 were actually calculated in 2002-3; the values published in 2011 were the result of a methodological extension and update to these earlier results, expressed in 2010 Euros. In the 2011 report, ICAO data (the latest available at the time) were used, in part, to update some of the earlier values to 2008 Euros, with subsequent estimates finally producing the values in 2010 Euros.

In this paper we present and extend this work, with a focus on the reactionary costs of delay and the implications for the design of new performance metrics. We have calculated the costs of delay to European airlines by four flight phases: at-gate, taxi, cruise extension and arrival management. Costs are assigned under three cost scenarios: ‘low’, ‘base’ and ‘high’. These scenarios are designed to cover the likely range of costs for European operators. The ‘base’ cost scenario is, to the greatest extent possible, designed to reflect the typical case. All calculations are undertaken for twelve aircraft: B733, B734, B735, B738, B752, A319, A320, A321, AT43, AT72, B744 and B763. The calculations use quantifications (by aircraft type) of block-hours (the time spent off-blocks - aircraft utilisation) and service-hours (the total time spent in service during the operational day).

The cost of delay is estimated separately for strategic delays (those accounted for in advance) and tactical delays (those incurred on the day of operations and not accounted for in advance). The type of strategic cost focused on is adding buffer to the airline schedule.

II. THE COST OF PRIMARY DELAY TO THE AIRLINE

A. Overview

In this paper, the focus is on the tactical cost of delay to the airline. It is important to note that these tactical costs are marginal (not unit) costs. They are thus cited for a range of delay durations (as shown in Table I), whereas strategic costs are scalable and may thus be cited per hour (not shown in this paper; see Cook and Tanner, 2011).

Fig. 1 shows which cost types are assessed at each level. Strategic costs and tactical costs are not independent: reactionary delays depend on the airline’s ability to recover from delay, due to the amount of schedule buffer, for example. If no buffers were used, the reactionary costs would increase markedly and the tactical costs would be significantly higher.

Reactionary delays are generally worse for longer primary delays and for primary delays that occur earlier in the operational day (when the knock-on effects in the network are greater). We will discuss the calculation of reactionary delay costs in Section III. We first summarise how the primary costs are estimated.

B. Fuel

The cost of fuel burned per minute is calculated for the three off-gate phases. Fuel costs doubled over the period 2004 - 2010. The base cost value in 2004 was 0.31 EUR/kg; the 2010 value used is 0.60 EUR/kg. The same fuel cost values are used for the strategic and tactical calculations. (A fuel carriage penalty is applied to arrival management, although these values are not used in this paper.)

C. Maintenance

Maintenance costs of delay incurred by aircraft relate to factors such as the mechanical attrition of aircraft waiting at gates (strategically or tactically) or aircraft accepting longer reroutes in order to obtain a better departure slot (tactically). The costs are based on values previously modelled in 2002, derived largely from interviews with eight European airlines, then updated to 2008 values using International Civil Aviation Organization (2008) data. The average European cost in 2008 was the same as the 2002 value. A small increase (5%) was then applied to produce 2010 values.

For the tactical values, marginal, time-based costs are derived from unit costs. Overheads are first removed and then a gate-to-gate model is used to apportion the maintenance cost between the airframe/components and powerplants across flight phases. The high intensity landing/take-off cycle maintenance costs (approximately 50%) are also excluded from the tactical calculations (the number of cycles is assumed to have been already fixed in the schedule). Both strategic and tactical at-gate costs are relatively low (compared to the other phases) because relatively little wear and tear on the airframe is experienced at-gate and the engines are off for the vast majority of this time.

D. Fleet

Fleet costs refer to the full cost of fleet financing, such as depreciation, rentals and leases of flight equipment. These costs are determined by service hours. Since utilisation has only a
very small effect on these costs, they are wholly allocated to
the strategic phase and the corresponding tactical delay costs
are thus taken to be zero.

Costs are based on values previously modelled in 2002
(sourced from airline interviews, literature and Airclaims data),
then updated to 2008 values using ICAO data (International
Civil Aviation Organization, 2008). The average European cost
fell by 15% from 2002 to 2008, although for several large
European airlines they fell by 50%, with further (smaller) falls
estimated for 2008 to 2010. The 2010 base scenario values are
20 – 35% lower than the 2002 values.

E. Crew

Pilots’ salaries generally increase by size of aircraft. Flight
attendants’ salaries are more consistent across all aircraft types.
In Europe, airlines typically pay crew fixed salaries,
supplemented by (relatively) small flying-time payments and
(cycles-based) allowances. Total cabin crew numbers are
driven by the maximum number of seats available (as in the
US).

Typical pilot and flight attendant salaries were calculated in
2008 for various European airlines, using their corresponding
payment schemes with realistic annual block/flight duty hours,
sectors flown and overnight stopovers. To update the 2008
costs to 2010 values, pay deals since 2008 for ten European
airlines were considered.

Tactically, cycles-based pay is subtracted from the annual,
total cost estimates such that the remaining proportion of the
salary is more accurately ‘time-based’. Airline on-costs are
included and overtime is considered.

Tactically, in certain cases, delays may not generate
additional crew costs, and the low cost scenario is set at zero
cost. The high cost scenario is based on overtime rates. The
base cost scenario is based on typical time-based costs. The
crew costs commonly apply to ground and airborne phases.

F. Passenger costs (to the airline)

Our cost estimations address airline delay costs – not wider
costs of delay (such as value of time costs internalised by
passengers and thus not impacting the airline). A cost of
passenger delay to the airline may be classified as either a
‘hard’ or ‘soft’ cost (see Fig. 2).

‘Hard’ costs are due to such factors as passenger rebooking,
compensation and care. Although potentially difficult to ascribe
on a flight-by-flight basis, due to accounting complications,
In a total cost model, primary costs need to be scaled up to the reactionary level. Reactionary delay may be further classified as 'rotational' and 'non-rotational'. Primary delays not only affect the initially delayed aircraft (flight 'X') on subsequent legs (rotational delay, e.g. flight 'Y'), but also other aircraft (non-rotational delay, flights ‘A’, ‘B’, … etc), as illustrated in Fig. 3. In our model, all reactionary delay is treated as at-gate delay, either for onward flights from the same airport (flights ‘Y’; ‘A’, ‘B’, … etc) or on subsequent rotations. We have built separate cost models for each of these.

Whereas in the calculations published in 2004, reactionary multipliers were applied differentiating by two types of delay (i.e. up to, and over, 15 minutes), the new model not only quantifies each reactionary delay as a function of the magnitude of the primary delay, but also more realistically assigns these costs over several rotations and applies caps to the rotational delays at costs comparable to those of cancelling a flight.

B. Rotational reactionary costs

a) Multipliers

The 2009 European reactionary to primary delay ratio of 0.82 is a system level ratio of departure delay as reported by airlines to EUROCONTROL's Central Office for Delay Analysis (EUROCONTROL Performance Review Unit, 2010). This means that for each minute of primary delay, on average, another 0.8 minutes (approximately) of reactionary delay are generated in the network. (The value was 0.85 in 2008, ibid.) This is often expressed in the literature as a multiplier, e.g. 1.8. For both 2008 and 2009, the ratio of rotational to non-rotational delay minutes was 88:12 (ibid.), as shown in Fig. 3. This is at the system level; it can vary significantly by airline.

Rather than simply multiplying all delay costs by a common factor (e.g. 1.8) to obtain the total network cost (primary plus reactionary cost), Beatty et al. (1998) studied delay propagation using American Airlines' schedule data, building delay trees, which included schedule buffers. Based (in part) on the Beatty model (ibid.), the multipliers we have developed also take into account the magnitude of the primary delay.

We first consider delayed flights by the delay ranges of Table I. This shows the proportion of Aircraft Communications, Addressing and Reporting System delays (for all delay causes) reported by airlines in 2009 (EUROCONTROL Performance Review Unit, 2010).

<table>
<thead>
<tr>
<th>Lower (mins)</th>
<th>1</th>
<th>5</th>
<th>15</th>
<th>30</th>
<th>60</th>
<th>90</th>
<th>120</th>
<th>180</th>
<th>240</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper (mins)</td>
<td>4</td>
<td>14</td>
<td>29</td>
<td>59</td>
<td>89</td>
<td>119</td>
<td>179</td>
<td>239</td>
<td>299</td>
<td></td>
</tr>
<tr>
<td>Proportion</td>
<td>.232</td>
<td>.367</td>
<td>.211</td>
<td>.119</td>
<td>.035</td>
<td>.015</td>
<td>.012</td>
<td>.004</td>
<td>.002</td>
<td>.003</td>
</tr>
</tbody>
</table>

To obtain the ‘basic’ multipliers for our model, the values of Beatty et al. (1998) are: (i) averaged across times of day; (ii) interpolated to match the ranges of Table I; (iii) adjusted such that when weighted by the proportion of delayed flights in each range of Table I, the average is 1.8 (the system level European multiplier). The results are shown in Table II. We call these ‘basic’ multipliers because we refine this method later in this paper.

<table>
<thead>
<tr>
<th>Primary delay (mins)</th>
<th>5</th>
<th>15</th>
<th>30</th>
<th>60</th>
<th>90</th>
<th>120</th>
<th>180</th>
<th>240</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic multiplier</td>
<td>1.49</td>
<td>1.67</td>
<td>1.94</td>
<td>2.47</td>
<td>3.01</td>
<td>3.54</td>
<td>4.61</td>
<td>5.67</td>
<td>6.74</td>
</tr>
</tbody>
</table>

Since the primary costs are non-linear, the basic multipliers of Table II will often overestimate the reactionary costs. Simply assigning, for example, one rotational delay of 60 minutes produces rather higher costs than two delays of 30 minutes. These multipliers need to be adjusted in different ways for narrowbody and widebody rotational delays. The joint constraints are: over how many rotations the reactionary delay may be distributed; how much delay is assigned to each rotation; and, avoiding night curfews on the final rotation. Model refinements are discussed next, first for narrowbodies.

b) Distributing the rotational costs over aircraft rotations

Focusing on Air France data for short delays, EUROCONTROL (2003) has compared, through modelling, propagated delay with additional local delay, flagging the importance of taking into account that a delay on one leg may cause the aircraft to miss its ATFM slot on a subsequent rotation, which may thus exacerbate the delay. As for delay recovery, peaks of propagated delay were not typically observed in the middle of the day (more chance of recovery during quieter periods) and the average departure delay from a given airport was typically less than the corresponding average arrival delay (recoveries may be made at-gate). The importance of load factors was also mentioned in this report: “the sensitivity analysis shows that a 45-minute scheduled station stop time does not allow a delay recovery if the departure plane is 80% loaded”.

Using US flight data for 2007, Kondo (2009) has examined propagation multipliers by tracking aircraft registrations. For the first propagated waves, delay multipliers were in a consistently narrow band, whereas they displayed quite large fluctuations for subsequent waves.

Although a number of research papers and reports have looked at the issue of delay propagation over several rotations, it is difficult to take conclusions forward to the generic European model that we are estimating here. We can, however, derive basic models from knowledge of narrowbody and widebody operations. In the next two sub-sections, we will derive a model for each, respectively.

1 Older US research (Boswell and Evans, 1997), using historical airline-reported delays for 1993, derived a rotational reactionary multiplier of approximately 1.8. More recently, Baden et al. (2006) reported that, on average, approximately one third of the US flight delays they modelled (for 2004 and 2000) was reactionary delay.
c) Refining the basic rotational model for narrowbodies

Narrowbodies typically have five rotations per day (Cook and Tanner, 2011). If the first rotation has a primary delay of around 2 hours, this causes another 4 hours in total of rotational delay (as a good approximation - interpolation from the Table II actually predicts that 1hr53 of primary delay produces, on average, 4hr00 of rotational reactionary delay, using the 88:12 ratio cited earlier). Let us consider these 4 hours.

Considering typical distributions of reactionary delay as percentage ratios over three waves, likely distributions are diminishing ones over subsequent rotations, such as 60:30:10 or 50:30:20, or those with interim recovery, such as 60:10:30.

For simplicity in the base case scenario, a simple 50:50 split (equally over two rotations) is used, as a reasonable approximation. This simple split only yields cost differences of up to 25% for the passenger hard costs, compared with either a 60:30:10 or 50:30:20 split. (The total percentage cost differences are less, since passenger hard costs are not the only component of the cost of delay.)

Considering the low cost scenario, these four hours of rotational delay could be ‘accommodated' within a typical narrowbody operational day as one hour on each rotation after the primary delay, for example. In general, splitting the reactionary delay into four equal parts produces a relatively low cost estimate. This example reflects some initial recovery (primary delay of two hours; next rotational delay one hour) and no further recovery. Airlines may sometimes put more schedule buffer early in the day to absorb delay.

Assigning a large, at-gate rotational delay (such as four hours) to just one of the subsequent rotations after the primary delay, means, by definition, that zero rotational delay is assigned to the other rotations. If such a single, large delay is all on an early rotation, attaining zero delay on later rotations may require a large amount of schedule buffer or slack time at-gate (an enforced stay due to airport slot or timetabling constraints, for example). Such recovery might also be achieved with an appropriate aircraft swap. This is the limiting case adopted for the high cost scenario. (Delay recovery at-gate is discussed below.)

Simple models suggest that total reactionary delay of much more than four hours are difficult to allocate to typical narrowbody operational days, without making a significant change, such as cancelling one or more rotations. Our model caps these costs at the cost of four hours of total rotational minutes under any given scenario.

For a base case scenario B735 (113 seats) this equals EUR 17 230. This compares to an approximate estimate in EUROCONTROL’s ‘Standard Inputs’ (EUROCONTROL, 2009) of the average cost of cancelling a 120-seat narrowbody flight, of EUR 16 000.

A summary of the way in which we allocate and cap rotational reactionary delay over the three cost scenarios is shown in Table III. In Fig. 4, the amount of rotational delay per rotation is plotted as a function of the primary delay.

<table>
<thead>
<tr>
<th>Reactionary delay</th>
<th>Low scenario</th>
<th>Base scenario</th>
<th>High scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 &lt; hrs ≤ 4</td>
<td>4 rotations</td>
<td>2 rotations</td>
<td>1 rotation</td>
</tr>
<tr>
<td>4 &lt; hrs (capped)</td>
<td>4 x 1 hr</td>
<td>2 x 2 hrs</td>
<td>1 x 4 hrs</td>
</tr>
</tbody>
</table>

The grey shaded triangle in Fig. 4 shows where the total rotational reactionary delay is less than the primary delay. For example, inside the triangle, 15 minutes of primary delay causes around 10 minutes of rotational reactionary delay. Outside the triangle, 60 minutes of primary delay produces just under 80 minutes of reactionary delay; where this is assigned equally over two rotations, for example in the base cost scenario, this implies 20 minutes' recovery on the first such rotation. Such recoveries are largely made through schedule buffer and slack time at-gate, and sometimes by achieving a faster turnaround at the gate. The high scenario curve (all the delay is here assigned to one rotation) shows the total reactionary delay for any scenario.

d) Refining the basic rotational model for widebodies

It is less straightforward to generically model widebody rotations, compared with the narrowbody case, not least due to the geographical scale and variability of such operations. For example, some eastbound flights may have quite long layovers in order that the westbound return rotation will land after a morning curfew (or at a more desirable time of day for passengers). Widebody aircraft are more likely to have overnight layovers than narrowbodies and to be able to make substantial airborne recoveries, for example due to favourable winds. Time zone differences may also lead to schedule bunching at certain times of the day. Due to considerably longer flight times, widebodies have fewer rotations in an operational day and there may be more opportunity to absorb delay on certain rotations (for example, those with particularly long layovers). This suggests spreading widebody reactionary delay over fewer rotations than the narrowbody case.

For the high cost scenario, as with narrowbodies, the reactionary delay is assigned to one rotation. In consideration of the low cost scenario, it is highly unlikely that widebody delay will persist over such a long timescale as four rotations (as used in the narrowbody low cost scenario). For the low cost widebody scenario, a 50:50 split is assumed over two

Figure 4. Narrowbody rotational reactionary delay.
reactionary rotations (which would typically be spread over more than one operational day).

For the base cost scenario, a ‘statistical’ cost, split over 1.5 rotations, is assigned (1.5 being the mid-point between the low and high cost scenarios). This is statistically equivalent to allocating 60 minutes of primary delay split separately as 45 and 15 minutes.

Compared to the narrowbody case, with fewer rotations over which to distribute delay but longer layovers in which to potentially reduce them, we have judgementally assigned a cost cap of five hours to the widebody case (in practice, such flights might be cancelled at longer or shorter delay). This refers to a limit of assigning five hours of total rotational minutes under any given scenario. For a base case scenario B744 (403 seats) this equals EUR 106 400. This compares to an approximate estimate in EUROCONTROL’s ‘Standard Inputs’ (ibid.) of the average cost of cancelling a 400-seat widebody flight, of EUR 75 000.

Table IV summarises the allocation of rotational reactionary delay and corresponding caps for the three cost scenarios. Fig. 5 shows the amount of rotational delay, per rotation, as a function of the primary delay.

<table>
<thead>
<tr>
<th>Reactionary delay</th>
<th>Low scenario</th>
<th>Base scenario</th>
<th>High scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 &lt; hrs ≤ 5</td>
<td>2 rotations</td>
<td>‘1.5’ rotations</td>
<td>1 rotation</td>
</tr>
<tr>
<td>5 &lt; hrs (capped)</td>
<td>2 x 2.5 hrs</td>
<td>4hr20 + 0hr40</td>
<td>1 x 5 hrs</td>
</tr>
</tbody>
</table>

![Widebody rotational reactionary delay](image)

As with the narrowbodies, the high scenario curve may be compared with the grey shaded triangle, which then represents the zone in which the total rotational reactionary delay is less than the primary delay. As with the narrowbodies, the high cost scenario curve is the steepest of the three. The base curve is drawn for the larger of the two reactionary delays, calculated according to the split described above.

e) Assigning costs to rotational reactionary delay

The rotational reactionary model is applied in a tailored way for each of the cost elements. As stated, all reactionary delay is assumed to occur at-gate. For fuel, the cost is therefore zero, since the at-gate model assumes that the auxiliary power unit and engines are off. The other costs are discussed in the following paragraphs. The narrowbody and widebody cost caps apply to all of these calculations.

For the passenger costing, each new rotation is assumed to have new passengers on-board. This means that the reactionary minutes are multiplied by the corresponding hard and soft tactical costs.

For the tactical maintenance costs: (i) where the rotational reactionary delay is less than the primary delay, no additional costs are added, nor are any cost savings calculated, as a result of the implied recovery; (ii) where the rotational reactionary delay is equal to the primary delay, no additional costs are added; (iii) where the rotational reactionary delay is greater than the primary delay, the cost associated with the additional reactionary delay is added once only (even on subsequent rotations with respect to schedule, this still relates to the same aircraft and must not be double-counted).

For the tactical crew costs, care needs to be taken to avoid double-counting and to take account of crew changes. Crew are assumed to remain the same for two narrowbody rotations but to change on each widebody rotation. For narrowbody crew, rotations are costed as paired rotations - a combined calculation over both rotations is made for one set of crew. This applies to a primary rotation and a following reactionary rotation and/or to subsequent, sequential reactionary rotations. On crew change (narrowbodies and widebodies) with a delayed flight, whether a cost is incurred depends on whether the new crew were advised to start their shift later (or a crew swap was carried out): high cost scenarios are fully costed; for base scenarios, half the cost per minute is used; for all low cost scenarios, crew are costed at zero - see Section II-E.

For narrowbody crew only, within a paired rotation (no crew change) if the delay on the second rotation is less than the one before it, the cost difference is recovered. In such cases, this means that a reactionary cost correction is subtracted from the primary cost (except for an ‘apparent’ recovery solely arising as the result of a cap). This means that at very low delay for narrowbodies (only for 5 minutes of primary delay for jets, up to 15 minutes for ATRs), the total delay cost may be less than the primary delay (by up to EUR 10 for base scenario, up to EUR 40 for high scenario). Otherwise, for two sequential reactionary rotations with delay, the cost is only considered once (and, again, set to zero in the low cost scenario).

f) Reflecting on the rotational model

Although few researchers have quantified reactionary delay models, AhmadBeygi et al. (2008) have constructed delay propagation trees using data sets from two US airlines, a large hub-and-spoke carrier and a point-to-point, low-fare carrier. Using data presented in that paper (ibid.), Fig. 6 plots the total reactionary delay calculated for these two carriers (dashed, grey curve) and compares these with the total (rotational) reactionary delay curves plotted in Fig. 4 (narrowbody) and Fig. 5 (widebody) for our models (upper and lower black curves, respectively). These authors calculated reactionary delay for up to 180 minutes of primary delay, beyond which cancellations or other recovery mechanisms were cited as coming into operation (although not modelled). Exploring
various other useful propagation metrics, saturation effects emerged at higher primary delay.

![Graph](image)

**Figure 6.** Comparing two reactionary delay models.

The two sets of curves are remarkably similar for primary delays of up to 60 minutes’ and reach similar reactionary totals by 180 minutes of primary delay.Whilst we have attempted to use reasonable assumptions, based on basic operational considerations, to generate a realistic set of reactionary scenarios for rotational delay, further modelling is required. This should use tail-specific data covering several European airlines and also include passenger connectivity data. Such improvements could also support a more refined consideration of turnaround recovery, drawing on the corresponding literature. For example, based on a Bayesian network model to predict future delays that incorporated interrelationships among delay causal factors, analysing the contribution of each phase to the arrival delay, Laskey et al. (2006) discuss turnaround recoveries in relation to preceding arrival delay. In Fricke and Schultz (2009), the potential for tactical absorption of delays during the turnaround process by means of dynamically scheduling buffer times is investigated.

C. Non-rotational reactionary costs

The non-rotational reactionary costs are more straightforward to allocate. It has already been stated that these represent 12% of all reactionary delays. Using 12% with the basic multipliers of Table II gives a simple estimate of the number of non-rotational reactionary minutes for each primary delay. It is assumed that these are all experienced by secondary aircraft waiting at-gate (flights ‘A’, ‘B’, … etc, Fig. 3) for the causal aircraft (flight ‘X’). No modelling of passenger or crew dependencies between the secondary and causal aircraft is included. Each non-rotational reactionary delay is thus treated as a new at-gate delay.

In terms of estimating typical aircraft connectivities in Europe, various approaches are possible. Analysing actual booking data from global distribution systems would be time consuming and prohibitively expensive, although this would give a very good estimate. Another method would be to examine permutations of origin and destination via all possible airports with agreed minimum connection times, using Official Airline Guide data. This would be a very large computational task, however. A simpler approach is to use the total number of flights for each of the twelve supported aircraft and to normalise these to produce a distribution which totals 100%. These flight frequencies cannot be guaranteed as a representative estimate of connection frequencies but this seems to be a reasonable approach. These aircraft represent over 50% of all IFR flights in 2009 (Central Flow Management Unit data, not shown; EUROCONTROL Performance Review Unit, 2010).

For each primary delay, the non-rotational reactionary minutes (primary delay x [1 – (basic multiplier)] x 12%) are converted to an at-gate cost for each of the twelve supported aircraft and then proportioned over the normalised distribution of these aircraft to give the weighted average. In the absence of superior data, the passenger costs are assigned (by judgement) to eight rotations, which could be considered as being spread over two rotations for each of four connecting aircraft. Whilst no specific cap is applied, this spreading of the cost reduces what would otherwise be an unrealistically high assignment of the passenger costs as single lots.

For the crew and maintenance costs, this cost spreading is not an issue, since the per-minute cost is not a function of the magnitude of the delay. These costs are assigned as a simple ratio (12%/88%) of the corresponding rotational reactionary cost. Where this corresponds to a negative crew cost (the correction applied at some lower delays for crew - see Section III-B(e)), zero cost is assigned.

IV. THE TOTAL COST OF DELAY

A. Summary

Passenger costs dominate at-gate delays (and hence reactionary costs), whilst fuel costs form a significant proportion of en-route delay costs at lower delay. In consideration of the importance of the passenger cost of delay to the airline, separate papers have been produced detailing the methodologies for the hard (Cook et al., 2009) and soft costs (Cook et al., 2011).

B. Breakdown of total cost by components

Fig. 7 shows the individual components (base cost scenarios are used) for a B738 and B744. Although fuel costs are a significant proportion of en-route costs at lower delay, as stated, they become proportionally less as the length of en-route delay increases. At higher delay, this proportion levels off (not shown) at 8% (B738) and 10% (B744). At 120 minutes, the en-route costs are dominated by the passenger costs (from 80-90% across all aircraft types).

The contribution of the reactionary cost varies by phase and delay duration and is an important part of the total cost consideration. For the B744 at-gate delay, it is the second highest cost (being slightly larger than the crew costs, as en route). This reflects the high cost of passenger delays for the widebody (compare the B738 reactionary costs).

---

2 Where the majority of delays are experienced in any case – see Table I. The average delay per flight in 2010 (across all causes) was 14.8 minutes (EUROCONTROL, 2011).

3 Funded by the same EUROCONTROL Research Grant scheme as this paper (see ‘Acknowledgement’).
C. Using aircraft weight to extend the cost model

In order to extend the cost model beyond the core twelve aircraft, we examine cost correlations with Maximum Take-Off Weights (MTOWs). These are obviously correlated with the size of the aircraft, and are also highly correlated (all $r^2 > 0.9$, core twelve aircraft, data not shown) with en-route fuel burn, passenger numbers and crew costs (the relationship between aircraft size and crew costs is raised in Section II-E).

MTOW not only correlates well with delay costs for all $t$ evaluated (see Table II), it is also already used to determine various charges in aviation, in accordance with ICAO principles of charging by weight (and distance, where appropriate). These principles are based on the assumption that the “value of the service generally increases as aircraft payload increases”, whereas using weight “less than proportionately” prevents the inequitable treatment of heavier aircraft, which may achieve greater productivity and efficiency (International Civil Aviation Organization 2007, 2009). All European en-route charges are based on the square root of MTOW (EUROCONTROL Central Route Charges Office, 2011).

The next stage of this work will be to establish suitable fits on the curves describing the total cost of delay as a function of delay duration, such as those shown in Fig. 9 for the B738 base cost scenario, at-gate and en-route. Despite good overall fits across the whole delay range (the quadratics shown both have $r^2 > 0.98$), the residuals at either low or high delay ($t > 300$ mins, where caps take effect) are unsatisfactory with a given, simple fit.

Solutions which suggest themselves are piecewise or logit functions. Another option is to focus on solutions at lower delay, for example up to 120 minutes, which would be satisfactory in the vast majority of delay recovery and/or flight prioritisation cases.

V. BEYOND CURRENT COST OPTIMISATION MODELS

A. Cost optimisation

Under future trajectory concepts, both “operator mission” and the “best business outcome” (Section I-A) should include the minimisation of costs to the airline and not just focus on delay minutes. Since the relationship between delay duration and cost is non-linear, optimising the former is not the same as optimising the latter.

It is important that the costs optimised are a true reflection of those actually incurred by the airline. Until such time as true, real-time costs are available (for example through dynamic cost indexing techniques, integrating flight planning and passenger reaccommodation tools), appropriate cost models provide valuable, interim insights. Such tactical delay cost models should: (i) be based on marginal costs; (ii) include the reactionary costs; (iii) reflect the cost non-linearities.
We have recently produced a paper reviewing some of the key modelling in the literature on such cost optimisation models, setting this in the context of flight prioritisation principles. Although most models in the literature are hypothesised cost functions, it has been shown that convex / super-linear cost functions generally provide superior solutions, for example with reduced bias towards certain types of operations. An appealing approach by Castelli et al. (2010) develops a novel mechanism for slot allocation based on market principles and enables airlines to pay for delay reduction or receive compensation for delay increase. These authors employ a non-linear and marginal, stochastic cost function derived from Cook et al. (2004).

An important consideration for the implementation of any economic mechanism is acceptability by the airline community. Using MTOWs has the advantage of being already historically established as a charging mechanism by ICAO in a number of contexts. It thus seems an appropriate and transparent factor for use in cost mitigation, in combination with the duration of the delay saving, through the non-linear (for example, quadratic) function described.

Industry reporting on delay is very much flight-centric. It lacks a formally adopted set of passenger-centric metrics. Sometimes, these types of metric may be aligned; in other cases, they may give contradictory results. For example, cancellations often increase flight punctuality statistics (by removing a potentially long delay from the dataset), whilst producing an extremely detrimental effect on passenger punctuality.

B. Passenger centricity

Consider these principles in the case of two delayed flights, ‘A’ and ‘B’, each operated by a B738 in Europe. Table V shows example (but reasonable) values for the associated revenue to the airline, the primary and total costs of delay to the airline operator (AO) for 30 minutes of primary delay, and the total passenger delay. The total cost of delay and the total passenger delay in Table V include these reactionary effects.

<table>
<thead>
<tr>
<th>Flight</th>
<th>Revenue</th>
<th>Primary delay</th>
<th>Primary AO delay</th>
<th>Total AO delay</th>
<th>Total passenger delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>€ 20 000</td>
<td>30 mins</td>
<td>€ 2 000</td>
<td>€ 2 200</td>
<td>120 hours</td>
</tr>
<tr>
<td>B</td>
<td>€ 10 000</td>
<td>30 mins</td>
<td>€ 1 500</td>
<td>€ 2 300</td>
<td>100 hours</td>
</tr>
</tbody>
</table>

This illustrates some of the challenges facing a quantified approach to flight prioritisation. How is it to be decided, under future trajectory management paradigms, which flight to prioritise under capacity constraints? Both have an equal primary delay of 30 minutes. Very unlikely to be an acceptable (or workable) principle with airlines, is prioritisation by some measure of ‘ability to pay’ (such as flight revenue), although this may be aligned with primary delay cost (favouring recovery of flight ‘A’). Taking reactionary costs into consideration, flight ‘B’ is the prioritisation candidate, a conclusion again reversed if based on total passenger delay incurred in the network.

Flight-centric and passenger-centric metrics have been discussed by Manley and Sherry (2008), who examine the trade-off between flight and passenger delay, and also between airline and passenger equity in US slot allocation. A model was developed to calculate metrics for efficiency (delay minutes) and equity (unbiased distribution of delay). Comparing different rationing rules, it was found that passenger delays could be significantly decreased with a slight increase in total flight delay. Flight cancellations were included in the model. Passenger numbers were calculated using available seats multiplied by averaged load factors (as in Section II-F), whereas 100% load factors were applied to international flights (unlike our model).

Compared to the current ration by schedule (RBS), rationing by aircraft size (three priority queues: ‘heavy’, ‘large’ and ‘small’ aircraft) decreased 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.

Using flight data to estimate passenger trip delay for 1 030 routes between the 35 busiest airports in the United States in 2006, Wang (2007) showed that passengers scheduled on cancelled flights or missed connections represent only 3% of total enplanements but generated 45% of total passenger trip delay. Findings were consistent with those of Calderón-Meza et al. (2008), who analysed US flight data for January through October 2007 on 5 224 routes between 309 airports, in that both estimates of average passenger trip delay for passengers on cancelled flights were 11 – 12 times greater than the average for passengers on ‘delayed’ flights (i.e. delayed by 15 minutes or more).

SESAR Workpackage E is currently exploring these issues in the ‘POEM’ project (Passenger-Oriented Enhanced Metrics). Recent (unpublished) results from the early design of metrics required to significantly progress the state of the art, applying complexity science techniques, have suggested five basic properties of such metrics. Aspects of these will be evaluated, in consultation with stakeholders, and prioritised for development. The properties are summarised as follows. Selected examples of the properties are given for illustration, although many dimensions have actually been identified already for each. (a) Mathematical construction - using complexity science methods, compared to traditional statistics. (b) Type of measure – counts (e.g. cancellations), measures of centrality, measures of dispersion (c) Temporal and spatial scales - time planning horizons; airports or networks. (d) Centricity - flight-centric, passenger-centric, node-centric. (e) Units of measurement - time, cost, distance inefficiencies.

---

4 Submitted to Journal of Aerospace Operations (please contact authors for updated details).

5 http://www.sesarju.eu/programme/workpackages/wpe

6 Delay variance itself causes costs to be incurred by the airline in terms of the extra schedule buffer consequently required. Both departure and arrival variances increased quite markedly in 2010 compared with 2009 (EUROCONTROL, 2011).
Such metrics will be explored using, and partly drive the design of, a dedicated European ATM model, with comprehensive passenger connectivity data (from IATA) and aircraft-tracked flight data (from EUROCONTROL), under different operational scenarios (such as new flight prioritisation principles). As evidenced by the literature, it is important that cancellations are modelled, which raises several challenges in the (current) absence of European data as robust as those available in the US.

Trade-offs will be quantified between the new metrics developed and existing metrics (such as average flight delay), and between equity and efficiency (Lulli and Odoni, 2007, have shown that fundamental conflicts may arise between equity and efficiency (minimising the cost function) in European ATM models).

Passenger-centric metrics have an increasingly valuable role to play in the development of the future management of delay and its propagation. In the context of 4D contract negotiations, and wider ATM, it is important to take account of contrasting flight- and passenger-centric effects, whilst mindful of the fact that costs to the airlines and passenger disutility will both continue to be non-linearly driven by delay relative to the original schedule.

ACKNOWLEDGMENT
The University of Westminster gratefully acknowledges the generous support of the European Organisation for the Safety of Air Navigation (EUROCONTROL) for enabling this research to be carried out, including a range of valuable technical and advisory inputs from EUROCONTROL’s Performance Review Unit. This paper has been co-financed by EUROCONTROL under its Research Grant scheme. The content of the work does not necessarily reflect the official position of EUROCONTROL on the matter. © 2011, EUROCONTROL and the named authors. All Rights reserved.

REFERENCES


Cook A., Tanner G. and Anderson S. (2004). Evaluating the true cost to airlines of one minute of airborne or ground delay, University of Westminster, for EUROCONTROL Performance Review Commission.


EUROCONTROL (2009). Standard Inputs for EUROCONTROL Cost Benefit Analyses (Ed. 4.0), EUROCONTROL Headquarters, Brussels.


EUROCONTROL Experimental Centre (2003). Flight delay propagation - synthesis of the study, M3 Systems, EEC Note No 18/03.

EUROCONTROL Performance Review Unit (2010). Personal communication.


