Stochastic Modelling of Aircraft Queues: A Review

Rob Shone, Kevin Glazebrook, Konstantinos Zografos

Lancaster University, Department of Management Science, Lancaster, United Kingdom
r.shone@lancaster.ac.uk, k.glazebrook@lancaster.ac.uk, k.zografos@lancaster.ac.uk

Abstract

In this paper we consider the modelling and optimal control of queues of aircraft waiting to use the runway(s) at airports, and present a review of the related literature. We discuss the formulation of aircraft queues as nonstationary queueing systems and examine the common assumptions made in the literature regarding the random distributions for inter-arrival and service times. These depend on various operational factors, including the expected level of precision in meeting pre-scheduled operation times and the inherent uncertainty in airport capacity due to weather and wind variations. We also discuss strategic and tactical methods for managing congestion at airports, including the use of slot controls, ground holding programs, runway configuration changes and aircraft sequencing policies.

Keywords: Aviation; Queueing systems; Stochastic modelling

1. Introduction

Many of the busiest airports around the world experience very high levels of traffic congestion for lengthy periods of time during their daily operations. This is due to a rapid growth in demand for air transport services, combined with physical and political constraints which usually prevent the expansion of airport infrastructure in the short-term. Congestion increases the likelihood of flights being delayed, and these delays may propagate throughout an airport network, with serious financial consequences for airlines, passengers and other stakeholders (Pyrgiotis et al., 2013). As airport slot coordinators and traffic controllers strive to improve the efficiency of their operations, there is considerable scope for new and innovative mathematical modelling techniques to offer valuable insight.

The capacity of the runway system represents the main bottleneck of operations at a busy airport (de Neufville and Odoni, 2013). When demand exceeds capacity, queues of aircraft form either in the sky (in the case of arriving aircraft) or on the ground (in the case of departures). The purpose of this paper is to present a concise review of the methods used by researchers to model aircraft queues since research in this area began in earnest about 60 years ago. Aviation in general is currently a very active research area, and our review will touch upon some of the wider topics that are closely related to aircraft queue modelling, including demand management strategies and the potential of strategic and tactical interventions to improve the utilisation of scarce resources at airports. Thus, we intend to discuss aircraft queues not only from a mathematical modelling perspective, but also in the context of the optimisation problems frequently posed in the literature.
Of course, queueing theory itself is also a vast topic and there is no common agreement on which of the classical models (if any) are most appropriate in the context of air traffic. Classical queueing theory texts such as Kleinrock (1975) tend to focus on models which are mathematically tractable, such as those with Markovian distributions for customer inter-arrival times and/or service times. Generally, closed-form “steady-state” expressions for expected queue lengths, waiting times and other performance measures are available only in cases where the parameters of these distributions are stationary and customer arrival rates do not exceed service rates. However, demand for runway use at a typical airport varies considerably during the day according to the schedule of operations, and runway throughput rates may be affected by weather conditions, sequencing rules and other factors. Moreover, demand rates may exceed capacity limits for extended periods of time at busy hub airports (Barnhart et al, 2003). We therefore need to consider time-dependent queues, for which steady-state results are of limited practical use (Schwarz et al, 2016).

In general, two of the most important characteristics of a queueing system are the customer arrival and customer service processes. We therefore organise this paper in such a way that these processes are discussed in Sections 2 and 3 respectively. Other, more application-specific aspects of modelling air traffic queues, including the effects of weather conditions and approaches for modelling airport networks, are discussed in Section 4. In Section 5 we provide a summary and discuss possible directions for future research.

2. Modelling demand for runway usage at airports

This section is divided into two parts. The first part focuses on the modelling assumptions often made regarding demand processes at airports, and the second part discusses related optimisation problems which frequently attract attention in the literature.

2.1. Modelling assumptions

Throughout this section we are concerned with the processes by which aircraft join queues waiting to use the runway(s) at airports. In the case of departing aircraft, these queues are located on the ground, usually at the threshold of the departure runway(s). Arriving aircraft, on the other hand, must wait in airborne “holding stacks” which are usually located near the terminal airspace, although in some cases they may also be “held” at other stages of their journeys by air traffic controllers (to control the flow of traffic into a congested air sector, for example). In many cases, a plane which lands at an airport will take off again (not necessarily from the same runway) within a couple of hours. This implies that the demand processes for arrivals and departures are not independent of each other, but in fact it is quite common in existing mathematical models for arrivals and departures to be treated as independent queues with time-varying demand rates which are configured according to the schedule of operations. The assumption of independence is undoubtedly an oversimplification, but it may not be particularly harmful if one considers a large airport with separate runways being used for arrivals and departures (this system is referred to as “segregated operations” and is used at London Heathrow, for example).
Nonhomogeneous Poisson processes (i.e. those with time-varying demand rates) were first used by Galliher and Wheeler (1958) to model the arrivals of landing aircraft at an airport. They used a discrete-time approach to compute probability distributions for queue lengths and waiting times. Subsequently, the Poisson assumption became very popular. Koopman (1972) considered the case of arrivals and departures sharing a single runway and modelled the Poisson arrival rates for both operation types as not only time-dependent but also state-dependent, with the two-dimensional state consisting of the queue lengths for arrivals and departures. This model allows for the possibility of “controlled” demand rates, whereby the demand placed on the system is reduced during peak congestion hours.

Hengsbach and Odoni (1975) extended Koopman’s approach to the case of multiple-runway airports, and claimed that the nonhomogeneous Poisson model was consistent with observed data from several major airports. Subsequently, Dunlay and Horonjeff (1976) and Willemain et al. (2004) used case studies to provide further evidence in support of the Poisson assumption. In the last few decades, nonhomogeneous Poisson processes have been widely adopted for queues of arrivals and departures at single airports (Kivestu, 1976; Bookbinder, 1986; Jung and Lee, 1989; Daniel, 1995; Hebert and Dietz, 1997; Fan, 2003; Mukherjee et al., 2005; Lovell et al., 2007; Stolletz, 2008; Jacquillat and Odoni, 2015a; Jacquillat et al., 2017) and also at networks of airports (Malone, 1995; Long et al., 1999; Long and Hasan, 2009; Pyrgiotis et al., 2013; Pyrgiotis and Odoni, 2016).

In case studies which rely on the nonhomogeneous Poisson model, the question arises as to how the demand rate functions for arrivals and departures – which we will denote here by \( \lambda_a(t) \) and \( \lambda_d(t) \) respectively – should be estimated. The schedule of operations for a single day at an airport can be used to aggregate the numbers of arrivals and departures expected to take place in contiguous time intervals of fixed length – for example, 15 minutes or one hour. The approach of Hengsbach and Odoni (1976) was to model \( \lambda_a(t) \) and \( \lambda_d(t) \) as piecewise linear functions, obtained by aggregating scheduled operations over each hour and then connecting the half-hour points using line segments, as shown in Figure 1. Various alternative data-driven methods can be devised. Jacquillat et al. (2017) modelled \( \lambda_a(t) \) and \( \lambda_d(t) \) as piecewise constant over 15-minute intervals, while Bookbinder (1986) used hourly data but relied on a three-point moving average method to remove “jump discontinuities” in the demand rates which would otherwise occur at the end of each hour.

Of course, airlines operate flights according to pre-defined schedules, so it is reasonable to question whether the Poisson assumption (which implies memoryless inter-arrival times) actually makes sense in this context. Various arguments can be put forward to make the case that, in practice, inter-arrival (and inter-departure) times are ‘sufficiently random’ for the Poisson model to be valid. For example, Pyrgiotis (2011) argues that large deviations from scheduled operations can occur as a result of flight cancellations, delays at “upstream” airports, gate delays for departures, variability of flight times due to weather and winds, etc. These deviations have the effect of “randomising” actual queue entry times.
Nevertheless, it is no surprise that various authors have challenged the Poisson assumption. In recent years, several authors have cited the development of the Next Generation Air Transportation System (NextGen) in the US as a possible reason for abandoning the Poisson model in the future. The NextGen system, which is expected to be fully in place by 2025, will allow four-dimensional trajectory-based operations (TBO). This should allow arrivals and departures to meet their scheduled operating times with greater precision (Joint Planning and Development Office, 2010). There is a similar ongoing project in Europe, known as Single European Sky ATM Research or SESAR (European Commission, 2014). In the light of these developments, there is considerable interest in modelling demand processes which have less variability than Poisson processes. Nikoleris and Hansen (2012) argued that the Poisson model cannot capture the effects of different levels of trajectory-based precision, because the variance in inter-arrival times is simply determined by the rate parameter. In a related piece of work, Hansen et al (2009) considered deterministic and exponentially-distributed inter-arrival times (both with time-varying rates) as two opposite extremes for the level of precision in meeting pre-scheduled operation times, and used case studies to show that the deterministic case could yield delay savings of up to 35%.

One type of demand process which has gained significant attention in recent years is the “pre-scheduled random arrivals” (PSRA) process. In PSRA queueing systems, customers have pre-scheduled arrival times but their actual arrival times vary according to random earliness/lateness distributions; for example, deviations from scheduled times may be normally or exponentially distributed. PSRA queues have been studied since the late 1950s (Winsten, 1959; Mercer, 1960), but their application to aircraft queues is a relatively recent development. An advantage of using the PSRA model is that variances of arrival and departure times can be controlled by choosing appropriate parameters for the earliness/lateness distributions, and this may be useful for modelling the more precise operation times expected under the NextGen and
SESAR systems. One disadvantage, however, is that PSRA queues are more difficult to study analytically, and indeed they are quite different from many of the classical models usually studied in queueing theory since inter-arrival times are neither independent nor identically distributed.

Guadagni et al (2011) made explicit comparisons between Poisson and PSRA demand processes and pointed out that PSRA queues exhibit negative autocorrelation, in the sense that time intervals which experience fewer arrivals than expected are likely to be followed by time periods with more arrivals than expected. Jouini and Benjaafar (2011) also made some progress in proving analytical properties of PSRA systems with heterogeneous customers and possible cancellations, although their model assumes that earliness/lateness distributions are bounded in such a way that customers are guaranteed to arrive in order of their scheduled times, which may not be suitable in an airport context. Caccavale et al (2014) used a PSRA model to study inbound traffic at Heathrow Airport, and argued that Poisson processes are a poor model for arrivals at a busy airport since, in practice, the arrivals stream is successively rearranged according to air traffic control (ATC) rules. Gwiggner and Nagaoka (2014) compared a PSRA model with a Poisson model using a case study based on Japanese air traffic, and found that the two models exhibited similar behaviour in systems with moderate congestion, but deviated from each other during high congestion. Lancia and Lulli (2017) studied the arrivals process at eight major European airports and found that a PSRA model with nonparametric, data-driven delay distributions provided a better fit for the observed data than a Poisson model.

Although time-dependent Poisson, deterministic and PSRA processes are by far the most popular choices for modelling aircraft queues found in the literature, a handful of other approaches have also been proposed. Krishnamoorthy et al (2009) considered “Markovian arrival processes” (MAPs), which generalise Poisson processes and can be studied using matrix analytic methods. Some authors have used observed data to fit nonparametric distributions for arrival and/or departure delays (Tu et al, 2008; Kim and Hansen, 2013). Finally, although our discussion throughout this section has focused on the use of time-dependent distributions, a number of authors have considered stationary demand processes (e.g. homogeneous Poisson processes) and attempted to gain insight by modelling aircraft as customers of different job classes (Rue and Rosenshine, 1985; Horonjeff and McKelvey, 1994; Bolender and Slater, 2000; Bauerle et al, 2007; Grunewald, 2016).

2.2. Optimisation Problems

Demand-related optimisation problems at airports are based on managing patterns of demand in such a way that the worst effects of congestion are mitigated, while at the same time the level of service provided (in terms of flight availability, punctuality, etc.) remains acceptable to passengers and other airspace users. Demand management strategies can be implemented at the strategic level, as part of an airport’s schedule design (which usually takes place several months in advance of operations) or at the tactical level, by making adjustments to aircraft flight plans in real time in order to prevent particular airports or airspace sectors from becoming heavily congested at certain times of day.
The busiest airports outside the US fall into the category of slot-controlled (level 3) airports, which means that airlines intending to use these airports for take-offs or landings must submit requests for time slots (typically 15 minutes long) during which they have permission to use the runways and other airport infrastructure. Although the US does not implement slot controls in the same manner, a small number of its airports are subject to the ‘high density rule’, which imposes hourly capacity limits (Madas and Zografos, 2006). Since slot allocation is usually carried out with a broad set of objectives in mind (including the need to design schedules which satisfy airlines’ requirements as equitably as possible), the resulting schedules do not always ensure effectively against the danger of severe operational (queueing) delays occurring in practice. For example, if too many flights are allocated to a small set of consecutive time slots, the consequences for airport congestion levels may be catastrophic. Thus, there is a need for demand management strategies to ensure that congestion mitigation is included as part of the slot allocation procedure.

Various authors (Barnhart et al, 2012; Swaroop et al, 2012; Zografos et al, 2012) have commented on the inherent trade-off that exists between schedule displacement and operational delays, as illustrated by Figure 2. At slot-controlled airports, certain time slots tend to be more sought-after by airlines than others. As a result, flight schedules which conform closely to airline requests are likely to result in large ‘peaks’ in demand at certain times of day. These schedules incur only a small amount of schedule displacement, since the requests from airlines are largely satisfied; however, severe operational delays are likely to be caused by the peaks in demand. Conversely, operational delays can be reduced by smoothing (or ‘flattening out’) the schedule to avoid such peaks, but this generally involves displacing flights to a greater extent from the times requested by airlines.

A useful survey of demand management strategies that have been implemented around the world is provided by Fan and Odoni (2002). These strategies can generally be divided into two categories: administrative and market-based. Administrative strategies involve setting ‘caps’ on the numbers of runway operations that can take place at an airport in a single time period, or a number of consecutive time periods. These ‘caps’ may apply to arrivals, departures or both, and are usually referred to in the aviation community as declared capacities (Zografos et al, 2017). The relevant optimisation problems involve deciding how these caps should be set optimally in order to ensure a satisfactory trade-off is achieved between congestion levels (which are usually modelled stochastically) and airlines’ operational needs (Swaroop et al, 2012; Churchill et al, 2012; Corolli, 2013). On the other hand, market-based strategies are based on using economic measures such as congestion pricing and slot auctions to relieve congestion during peak periods (Andreattta and Odoni, 2003; Fan, 2003; Pels and Verhoef, 2004; Mukherjee et al, 2005; Ball et al, 2006; Andreattta and Lulli, 2009; Pellegrini et al, 2012). A number of authors have directly compared administrative and market-based strategies using analyses and/or case studies (Brueckner, 2009; Basso and Zhang, 2010; Czerny, 2010; Gillen et al, 2016).
As mentioned earlier, demand management can also be done at a tactical level. Ground-holding programs can be used to delay departing aircraft in order to ensure that they do not arrive at their destination airports during periods of high congestion. This not only relieves congestion at busy airports, but also has the benefit of preventing aircraft from wasting too much fuel by being forced to wait in airborne holding stacks. In the broader realm of air traffic flow management (ATFM), aircraft can be directed by air traffic controllers to delay their arrivals at congested airports or air sectors by adjusting their speeds or routes. Optimisation problems related to ground-holding programs and ATFM operations have been well-studied in the literature. These problems usually do not involve stochastic queue modelling, but they do commonly take account of the uncertainty caused by weather conditions, enroute congestion etc. by incorporating probabilistic “capacity profiles” for destination airports and incorporating the probabilities for different capacity scenarios within optimisation models such as integer linear programs (ILPs). Some notable references include Terrab and Odoni (1992), Richetta and Odoni (1993), Vranas et al. (1994), Richetta (1995), Bertsimas and Stock (1998), Hall (1999), Ball et al. (2003), Inniss and Ball (2004), Kotnyek and Richetta (2006), Lulli and Odoni (2007), Mukherjee and Hansen (2007), Liu et al. (2008), Balakrishnan and Chandran (2014).
3. Modelling capacity and runway throughput at airports

This section is organised in a similar way to Section 2. The first part focuses on modelling assumptions related to capacity and runway throughput rates at airports, and the second part discusses some relevant optimisation problems.

3.1. Modelling assumptions

An airport’s capacity can be defined as the expected number of runway movements (either arrivals or departures) that can be operated per unit time under conditions of continuous demand (de Neufville and Odoni, 2013). It is very important to estimate an airport’s capacity accurately, since long queues of aircraft waiting to use the runway(s) may form as a result of imbalances between demand and capacity, and therefore capacity modelling must be used to inform the demand management strategies discussed in Section 2.2. However, an airport’s capacity is not simply a fixed quantity, but instead is time-varying and depends on a number of factors. Adverse weather conditions might increase the separation requirements between consecutive arriving aircraft, while strong winds may prevent certain runways from being used. Additionally, runway movements might be restricted at certain times of day due to noise considerations. For example, at Heathrow Airport, the period between 11:30pm and 6:00am is known as the “Night Quota Period”, with traffic restrictions imposed by the Department for Transport (Heathrow Airport, 2018).

Blumstein (1959) produced a seminal paper in which he explained how to calculate the landing capacity of a single runway (i.e. when it is used for arrivals only) based on aircraft speeds and separation requirements. Hockaday and Kanafani (1974) generalised Blumstein’s work by deriving expressions for the capacity of a single runway under three different modes of operation: arrivals only, departures only and mixed operations. Newell (1979) showed how to extend these analyses to airports with multiple runways under various different configurations. A key principle which emerged from these early contributions was the importance of taking into account different possible fleet mixes and sequencing strategies. When one runway movement is followed by another, the movements in question are subject to a minimum time separation which depends not only on the types of movements involved (arrivals or departures), but also on the types of aircraft. To be more specific, aircraft can be categorised into different ‘weight classes’. Heavy aircraft generate a lot of wake turbulence, which can be dangerous to lighter aircraft following too closely behind (Newell, 1979). Therefore, in airport capacity calculations, one must take into account the relative expected frequencies of different ‘weight pairs’ (e.g. heavy-light, heavy-heavy, etc.) and use these to calculate average time separations between movements.

Gilbo (1993) developed the idea of the runway capacity curve (referred to by subsequent authors as a “capacity envelope”), as shown in Figure 3. This curve represents the departure capacity of an airport as a convex, nonincreasing function of the arrival capacity. The shape of the curve depends on various time-varying factors, including weather conditions, the runway configuration in use and the aircraft fleet mix. However, the essential principle is that each point on the capacity envelope represents a feasible pair of capacity values for arrivals and
departures during the time period for which the envelope applies. Various authors have provided detailed descriptions of how airport capacity envelopes can be constructed using both empirical and analytical methods (Lee et al., 1997; Stamatopoulos et al., 2004; Simaiakis, 2013) and these capacity envelopes have been incorporated into various types of optimisation problems, which we discuss further in Section 3.2.

The capacity of an airport is naturally related to the concept of a service rate in queueing theory, since it specifies how many runway movements (which we can think of as ‘services’ of aircraft) can be achieved in a given time interval. Several early studies modelled the queueing dynamics at airports using nonstationary deterministic models, with the arrival and service rates defined according to flight schedules and capacity estimates respectively (Kivestu, 1976; Hubbard, 1978; Newell, 1979). However, at the same time, interest was developing in modelling aircraft service times stochastically. Koopman (1972) proposed that the queueing dynamics of an airport with a nonhomogeneous Poisson process for arrivals and \( k \) runways (modelled as independent servers) could be bounded by the characteristics of the \( M(t)/D(t)/k \) and \( M(t)/M(t)/k \) queueing systems. The former system – in which the service process is nonstationary and deterministic – can be regarded as a “best-case” scenario, since queueing delays are shorter in the case of predictable service times. The latter system – with exponentially-distributed service times – is a “worst-case” scenario, in which highly variable service times cause average queueing delays to increase. Koopman used numerical solution of the Chapman-Kolmogorov equations (assuming a finite queue capacity) to estimate queue length probability distributions.

![Figure 3: A piecewise linear capacity envelope for a particular time interval, adapted from Stamatopoulos et al. (2004). Each point on the envelope represents a feasible pair of capacity values. Points 1 and 4 represent “all arrivals” and “all departures” policies respectively. Point 2 represents a sequencing strategy where departures are freely inserted during large inter-arrival gaps, and Point 3 is a “mixed operations” point.](image-url)
Kivestu (1976) proposed an $M(t)/E_k(t)/s$ queueing model for aircraft queues, in which the service time distribution is Erlang with $k$ exponentially-distributed service phases. This approach is closely related to that of Koopman (1972), since the cases $k = 1$ and $k = \infty$ represent exponential and deterministic service times respectively. However, Kivestu also introduced a fast, practical numerical approximation method for the time-dependent queue length probabilities in an $M(t)/E_k(t)/1$ queue, which became known as the DELAYS algorithm. Subsequently, DELAYS – as well as the $M(t)/E_k(t)/1$ model for aircraft queues itself – has become very popular, and has been used for estimating queueing delays in a variety of settings (Abundo, 1990; Malone, 1995; Fan and Odoni, 2002; Stamatopoulos et al, 2004; Mukherjee et al, 2005; Lovell et al, 2007; Churchill et al, 2008; Hansen et al, 2009; Pyrgiotis and Odoni, 2016). An advantage of using DELAYS is that it estimates the full probability distribution for the queue length at any given time. This is useful because, in practice, airports are often interested in tail-based performance measures such as the expected number of queueing delays that will exceed a given threshold.

The $M(t)/E_k(t)/1$ model for aircraft queues can be regarded as somewhat macroscopic, since it does not explicitly take into account fleet mixes and separation requirements between different aircraft types. Instead, it assumes that such considerations are implicitly accounted for via the use of an Erlang distribution for service times, whose variance can be controlled by adjusting the parameter $k$ (with larger values implying less variance). Models which explicitly consider runway occupancy times for different classes of aircraft have been proposed by a number of authors. Hockaday and Kanafani (1974) and Stamatopoulos et al (2004) modelled these using normal distributions, while Jeddi et al (2006) suggested beta distributions and Nikoleris and Hansen (2015) used Gumbel random variables. In models where different aircraft types are considered explicitly, there is certainly some justification for using service time distributions with very small variances – or even deterministic service times – since the time lapse between two consecutive aircraft entering the runway depends on separation guidelines which are enforced by air traffic controllers according to the aircraft weight classes. Indeed, a number of authors have used stochastic queueing formulations in which service times are deterministic, such as $M(t)/D(t)/1$ (Galliher and Wheeler, 1958; Daniel, 1995; Daniel and Pahwa, 2000) or $PSRA/D/1$ (Caccavale et al, 2014; Gwiggner and Nagaoka, 2014). The use of trajectory-based operations in the future (as discussed in Section 2.1) may provide further justification for considering deterministic service times.

In traditional queueing system formulations, the description of the service process includes not only the service time distribution but also the number of servers, finite queue capacity (if applicable) and the queue discipline. We therefore address the relevant modelling assumptions here in an aviation context. The assumption of a single server (as in the $M(t)/E_k(t)/1$ model, for example) is surprisingly common in the literature, even when the airport being modelled has more than one runway. One possible explanation for this is that even when an airport has multiple runways, there is usually some inter-dependence between them, which implies that it is inappropriate to model them as independent servers (Jacquillat, 2012). For example, runways may intersect each other – or even if they do not, they may be too closely-spaced to allow independent operations, since the effects of wake turbulence may create “diagonal separation...
requirements” between aircraft on different runways (Stamatopoulos et al., 2004). Nevertheless, the single-server assumption – which effectively models the runway system as a ‘black box’ processing arrivals and departures – is undoubtedly a simplification which, arguably, has been over-used in the literature.

A finite queue capacity is usually not considered an essential component of an airport queueing model, since in practice it is rare for aircraft to be denied access to an airport due to over-congestion (this would be referred to as ‘balking’ in queueing theory). Nevertheless, it should be noted that certain numerical methods for estimating queue length probability distributions, including the DELAYS algorithm and numerical solution of the Chapman-Kolmogorov equations, must assume a finite queue capacity for computational purposes. In practice, the queue capacity used in these methods is chosen to be large enough to ensure that it has very little impact on estimates of performance measures.

Finally, models which assume independent queues for arrivals and departures use the first-come-first-served queue discipline (FCFS) almost universally, unless they are intended to examine the effects of different sequencing policies. The FCFS assumption is largely consistent with ATC procedures in practice (Pyrgiotis, 2011). Some authors, however, have considered priority queues in which arrivals are given priority for service over departures (Roth, 1979; Horonjeff and McKelvey, 1994; Grunewald, 2016).

3.2 Optimisation problems

Optimisation problems related to service rates in aircraft queues may involve the strategic or tactical control of runway configurations, the dynamic balancing of service rates between arrivals and departures, the sequencing of aircraft using the runway(s) or some combination of these. In this subsection we provide examples from the literature.

As mentioned in Section 3.1, the shape of the airport capacity envelope (see Figure 3) depends on a number of operational factors which may vary during a day of operations. One such factor is the runway configuration. Airports with multiple runways may control which ones are active at any given time, although sometimes this choice is constrained by wind conditions which make it unsafe for aircraft to take off or land in a particular direction (Jacquillat and Odoni, 2015a). Ramanujam and Balakrishnan (2015) used empirical data to analyse the runway configuration selection process at US airports and aimed to predict the configurations chosen under different wind, weather and demand conditions. As discussed earlier, any point on the capacity envelope associated with a particular runway configuration represents a pair of attainable capacity values for arrivals and departures. It is natural to interpret capacity values as service rates which can be incorporated within queueing models. Various authors have considered optimisation problems in which an airport capacity envelope (or sequence of envelopes) is given, and the objective is to choose a sequence of points (i.e. service rate pairs) on these envelopes which will optimise a performance measure related to queue lengths or flight operation times (Gilbo, 1993; Gilbo, 1997; Hall, 1999; Dell’Olmo and Lulli, 2003). Other authors have extended this approach by modelling the runway configuration as a decision variable, so that the decision-maker must jointly optimise runway configurations and
arrival/departure service rates (Li and Clarke, 2010; Weld et al, 2010; Bertsimas et al, 2011). Jacquillat et al (2017) (see also Jacquillat and Odoni, 2015b) also considered a similar problem, but made an important contribution by including stochastic queueing dynamics (based on an $M(t)/E_k(t)/1$ formulation) in their model. Prior to this, deterministic queueing dynamics had generally been assumed for such problems, with solutions found using ILPs.

Aircraft sequencing (also known as runway scheduling) problems involve planning the order that arriving and/or departing aircraft will use the runway(s) in such a way that a certain performance measure is optimised. As discussed earlier, the time separations between consecutive runway movements depend on the types of aircraft involved, and significant amounts of time can be lost if smaller aircraft often have to follow heavier ones. Throughput rates will generally be maximised if groups of similar aircraft are allowed to take off or land consecutively, but other constraints and objectives must also be taken into account. For example, individual aircraft might have to take off or land within fixed time windows, and the objective(s) might include allowing aircraft to take off (land) as close as possible to pre-specified ‘preferred’ take-off (landing) times. “Constrained position shifting” (CPS), whereby an aircraft’s position in the sequence is allowed to deviate by only a certain maximum number of places from its position in a “first-come-first-served” sequence, is another common way of enforcing constraints (Dear, 1976). “Static” aircraft sequencing problems are those in which the sequence of runway movements is optimised only once, and does not change in response to any subsequent events (Psaraftis, 1978; Beasley et al, 2000; Artioucheine et al, 2008; Balakrishnan and Chandran, 2010). On the other hand, in the dynamic version of the problem, the sequence is re-optimised every time new aircraft enter the terminal control area and become available for sequencing (Dear, 1976; Beasley et al, 2004; Murca and Muller, 2015; Bennell et al, 2017). Both versions of the problem are usually formulated as deterministic optimisation problems and solution approaches may include dynamic programming, branch-and-bound methods and metaheuristics (Potts et al, 2009). There has also been some interest in formulating stochastic runway scheduling problems. In these problems, the random variables may include the arrival times of aircraft in terminal areas, pushback delays for departures and taxiway times. Two-stage stochastic optimisation has been employed as a solution method (Anagnostakis and Clarke, 2003; Solveling et al, 2011; Solak et al, 2018).

4. Other modelling considerations at airports

In this section we discuss certain other aspects of modelling airport operations which have been touched upon only briefly in the previous sections.

Firstly, we address the subject of weather. One of the most obvious reasons for using stochastic (as opposed to deterministic) models for airport operations is the fact that weather and wind conditions can never be anticipated with complete confidence. Poor weather conditions cause visibility problems which can increase the separation requirements between consecutive runway operations and runway occupancy times, thereby effectively reducing airport capacities. Gilbo (1993) described how empirical data could be used to construct separate capacity envelopes for different weather categories. Other authors (Simaiakis, 2013; Jacquillat and Odoni, 2015a) have noted that, in practice, a distinction exists between “Visual
Meteorological Conditions” (VMC) and “Instrumental Meteorological Conditions” (IMC), which indicate “good” and “poor” weather respectively. Based on this distinction, VMC and IMC envelopes can be constructed for each possible runway configuration, with the IMC envelopes being smaller than the VMC ones but similar in shape.

Of course, knowing how to estimate airport capacity envelopes under different weather conditions is one thing, but simulating random weather changes within decision problems is quite another. When designing stochastic models for weather evolution, it makes sense to consult historical data in order to estimate the relative frequencies for different weather states. Modelling the random transitions between weather states can be done in different ways. Jacquillat and Odoni (2015a) used a nonstationary two-state Markov chain, with the transition probabilities from state “VMC” to “IMC” and vice versa estimated using historical data (see Figure 4). Other authors have opted for a semi-Markov model, in which the time spent in a particular weather state has a non-exponential distribution (Abundo, 1990; Peterson et al., 1995). In the literature on ground-holding problems discussed in Section 2.2, it is common practice to represent an airport’s capacity profile probabilistically by specifying probabilities for different weather scenarios (see, for example, Richetta and Odoni, 1993). Liu et al (2008) and Buxi and Hansen (2011) have discussed the use of clustering techniques for generating probabilistic capacity profiles. In problems where the decision-maker has the ability to switch between different runway configurations, it is also important to note that wind conditions may prevent certain configurations from being used. Jacquillat et al (2017) described the use of a Markov chain model for transitions between 13 different wind states in a case study based on JFK Airport in New York.

![Figure 4: A nonstationary two-state Markov chain model for weather evolution. The parameters $p_t, q_t \in [0,1]$ depend on the discrete time interval $t$.](image)

The previous sections of this paper have focused mainly on the modelling of aircraft queues at single airports. However, research has also been done into modelling airport networks. The relevant papers tend to focus on the propagation of delays around a network, referred to as the “ripple effect”. Long et al (1999) (see also Long and Hasan, 2009) developed the “LMINET model”, in which airports are modelled as a network of $M(t)/E_k(t)/1$ queues. Pyrgiotis et al (2013) developed the “Approximate Network Delays” (AND) model (first conceptualised in Malone (1995)), which iterates between a network queueing engine and a delay propagation algorithm for modelling network delays. The queueing engine is based on a network of
queues and relies upon the DELAYS algorithm discussed in Section 3.1, while the delay propagation algorithm explicitly considers individual aircraft itineraries. Baspinar et al (2016) used a similar model to investigate the effects of local disturbances (e.g. strike action or severe weather) at European airports. Czerny (2010) compared slot constraints with congestion pricing as alternative methods for managing demand in a network of airports, while Vaze and Barnhart (2012) used the AND model to test the effects of demand management strategies under different capacity scenarios. Campanelli et al (2016) discussed the use of agent-based simulations for modelling network delays. It should also be mentioned that all of the literature on ground holding problems (see Section 2.2) is network-related, since the decision to delay an aircraft’s departure from one airport is made with the intention of improving congestion at another. However, several of these papers consider simplified “star-shaped” networks in which a single “hub” airport is assumed to be the only one prone to congestion, and ground holding decisions made at other airports are based entirely on managing congestion at the hub airport.

Finally, some interesting papers have arisen from considering the differences in demand management and ATFM strategies at US and European airports. Odoni et al (2011) compared the demand-to-capacity relationships at Frankfurt International Airport (which is a slot-coordinated airport) and Newark Liberty Airport in New York. Frankfurt is subject to much stricter demand regulation than Newark, and consequently it performs better with respect to average flight delays, but the paper suggests that the economic benefits of increasing slot limits may outweigh the costs of increased delays. Swaroop et al (2012) investigated the slot controls in use at the four slot-controlled airports in the US and found that the costs of airport congestion were too high to justify the relatively relaxed slot constraints. Both of the aforementioned papers support the general view that slot controls in Europe are too strict, whilst in the US they tend to be too liberal. Campanelli et al (2016) investigated the differences in network delays between the US and European air traffic systems which are caused by different aircraft sequencing strategies.

5. Summary

Methods for modelling aircraft queues are continuously evolving. Nonstationary models based on classical queueing theory are still employed frequently. For example, the $M(t)/E_k(t)/1$ model continues to attract a lot of attention. With the ongoing development of systems based on trajectory-based operations (in particular, the NextGen system in the US and SESAR in Europe), we anticipate that models which allow the variances of inter-arrival times and/or service times to be controlled at a finer level – through the use of queue entry times based on pre-scheduled random arrivals (PSRA), for example – are likely to become more popular. In addition, we suggest that some of the simplifying assumptions that have been adopted almost universally over the last few decades – such as the single-server assumption for multiple-runway airports and the independence of queues for arrivals and departures – are likely to be relaxed as researchers increasingly aim to incorporate high-fidelity models of airport operations into their optimisation procedures.
This paper has touched upon some of the broader issues related to the stochastic modelling of aircraft queues, including demand management strategies and the tactical control of aircraft take-off and landing sequences. It is clear from our discussion that, in reality, the queueing dynamics at airports are influenced by a diverse range of factors, including the decisions made at different points in time by multiple stakeholders. From a strategic point of view, the decisions made regarding slot controls (or congestion pricing) at airports and the slot requests submitted by airlines are important for determining, several months in advance, the daily demand profiles at airports. However, the tactical decisions made by airports and air traffic controllers in ‘real time’ – which may be related to sequencing patterns, ground holding delays and runway configuration changes – are also critical for managing congestion. We conclude that the modelling and optimisation of queues and congestion levels at airports is a complex task which should be informed by field analyses and engagement with industry practitioners in order to maximise research impact.

Acknowledgements

This work has been supported by the Engineering and Physical Sciences Research Council (EPSRC) through Programme Grant EP/MO20258/1 “Mathematical models and algorithms for allocating scarce airport resources (OR-MASTER)”.

References


Airports. Proc. 7th USA/Europe Air Traffic Management R&D Seminar, Barcelona, Spain.


Shone, Glazebrook and Zografos


Shone, Glazebrook and Zografos


