

Scheduling Airline Reserve Crew using a Probabilistic Crew Absence and Recovery Model

Christopher Bayliss
IN3 – Computer Science Dept
Universitat Oberta de Catalunya
Spain, cbayliss@uoc.edu

Geert De Maere
Jason A. D. Atkin
ASAP
School of Computer Science
University of Nottingham, UK
gdm,jaa@cs.nott.ac.uk

Marc Paelinck
ITCDOR
KLM Royal Dutch Airlines
KLM Headquarters
The Netherlands
marc.paelinck@klm.com

Abstract

Airlines require reserve crew to replace delayed or absent crew, with the aim of preventing consequent flight cancellations. A reserve crew schedule specifies the duty periods for which different reserve crew will be on standby to replace any absent crew. For both legal and health-and-safety reasons the reserve crew's duty period is limited, so it is vital that these reserve crew are available at the right times, when they are most likely to be needed and will be most effective. Scheduling a reserve crew unnecessarily, or earlier than needed, wastes reserve crew capacity. Scheduling a reserve crew too late means either an unrecoverable cancellation or a delay waiting for the reserve crew to be available. Determining when to schedule these crew can be a complex problem, since one crew member could potentially cover a vacancy on any one of a number of different flights, and flights interact with each other, so a delay or cancellation for one flight can affect a number of later flights. This work develops an enhanced mathematical model for assessing the impact of any given reserve crew schedule, in terms of reduced total expected cancellations and any resultant reserve induced delays, whilst taking all of the available information into account, including the schedule structure and interactions between flights, the uncertainties involved, and the potential for multiple crew absences on a single flight. The interactions between flights have traditionally made it very hard to predict the effects of cancellations or delays, and hence to predict when best to allocate reserve crew and lengthy simulation runs have traditionally been used to make these predictions. This work is motivated by the airline industry's need for improved mathematical models to replace the time-consuming simulation-based approaches. The improved predictive probabilistic model which is introduced here is shown to produce results that match a simulation model to a high degree of accuracy, in a much shorter time, making it an effective and accurate surrogate for simulation. The modelling of the problem also provides insights into the complexity of the problem that a purely simulation based approach would miss. The increased speed enables potential deployment within a real time decision support context, comparing alternative recovery decisions as disruptions occur. To illustrate this, the model is used in this paper as a fitness function in meta-heuristics algorithms to generate disruption minimising reserve crew schedules for a real airline schedule. These are shown to be of a high quality, demonstrating the effectiveness and reliability of the proposed approach.

Keywords: Airline scheduling, reserve crew, crew absence, uncertainty, probabilistic model

1 Introduction

Airlines operate in an uncertain environment due to the effects of weather, congestion, crew unavailability and unscheduled aircraft maintenance. Airlines also operate in a competitive business environment, competing to provide customers with a desirable product (the schedule), whilst trying to minimise expenditure. Economic pressure pushes airlines towards schedules that maximise resource utilisation and have minimal slack. Such tight schedules are susceptible to delay propagation and cancellations when there is crew or aircraft unavailability. In addition, flights are often interlinked. For example an aircraft may arrive at a hub airport with one crew team, then take off for a later flight with another one, and the first crew may then join a different aircraft, so that any delays to the previous flight's arrival may also have an effect on this other flight. Predicting a delay for a flight may require consideration of a number of other flights, each of which may have a different predicted delay distribution and slack.

Airlines have limited capacity to recover from disruptions caused by delay or resource unavailability. One method of recovery is to utilise reserve crew—alternative crew members who will be available to cover for other crew who are absent. The regular crew schedule is the input for reserve crew scheduling. Reserve crew scheduling could be viewed as the process of patching up the operational robustness of the regular crew schedule. An improved allocation of the available reserve crew to standby duty periods is an attractive route to increased operational robustness, however the question is how to determine the number of reserve crew that should be available at different times.

The traditional way to determine the benefits from a specific allocation is to run Monte-Carlo simulations using delay distributions, but this can be very time consuming. An emerging direction of research has been the use of alternative mathematical approaches, which are becoming more common in recent years. The accuracy of these models for evaluating the effects of different reserve crew allocations has a huge effect upon the validity of the decisions which are made by utilising them. In contrast, this paper helps to solve the problem of scheduling reserve crew to minimise cancellations due to crew unavailability, by providing an improved method for evaluating the likely effects of different allocations of reserve crew (a descriptive model of the problem), and evaluating the benefits of these improvements within a (prescriptive) method for actually scheduling the crew.

This work focusses on the off-line problem of scheduling reserve crew duty times before the day of operation, before any information about day of operation uncertainties (crew absence and delays) is available. Our descriptive model evaluates candidate reserve crew schedules using a propagating-probabilities ap-

proach. In contrast to the off-line reserve crew scheduling problem, the on-line reserve use problem is concerned with decisions regarding the use of already scheduled reserve crew on the day of operation in response to disruptions such as absent crew and flight delays. In Sections 3 and 6 we discuss how the proposed model can also be used in an on-line context to evaluate alternative reserve use decisions such as using reserve crew or holding them back for larger disruptions later on for which they may be better utilised.

This work introduces an improved formulation of a model which was introduced by Bayliss et al. (2012) and featured a probabilistic model of crew absence and reserve crew utilisation. The improved model introduces a consideration of any delays which the use of reserve crew may actually introduce (in exchange for a cancellation) due to having to wait for reserves to be available (see Section 4.6).

In Bayliss et al. (2012), the inputs to the model were the probabilities of crew absence for a sequence of departures from an airport at which a number of reserve crew are based. These crew absence probabilities were then used to evaluate the effects that a given reserve crew schedule would have upon the probabilities of flight cancellations due to crew absence. The model was used as a surrogate objective function to provide a single valued measure of the quality of reserve crew schedules and is much faster than evaluation by simulation. The model itself is descriptive, in that it determines the likely effects rather than suggesting how to change the effects. It was used within a prescriptive system in Bayliss et al. (2012), where meta-heuristic algorithms utilised it in a search for reserve crew schedules that would minimise the expected cancellations due to crew absence, and that approach is continued here for evaluating the improved model.

This current work also addresses the question of when to schedule reserve crew, providing an improved model which can be utilised in the same way, but also addresses the simplifying assumptions which were discussed in Bayliss et al. (2012). The improved model is validated in terms of cancellation prediction accuracy by comparing the results against those of a large (and time consuming) simulation. The model is also used as the evaluation function in an off-line search through reserve crew schedules. The results will show that the validity of the descriptive model of the problem has been improved, and that this has enabled improvements in the (prescriptive) decision making model, resulting in allocation decisions which reduced the delays/cancellations.

This work provides: 1) a focussed investigation of an important problem faced by airlines; 2) a model which accurately predicts cancellation probabilities as a function of any given reserve crew schedule; 3) a validation of the model in terms of improved prediction accuracy and improved benefits from its use within a reserve crew scheduling application when applied to a realistic test instance.

The main motivations and contributions of this work are summarised as follows: 1) A previous probabilistic (descriptive) model of the impact of any given reserve crew schedule on the probabilities of flight cancellations due to crew absence has been extended and enhanced to increase its accuracy and applicability; 2) The simplifying assumptions made previously have been replaced with real world information; 3) The new model required the development of a tree generation algorithm for enumerating all of the sensible combinations of reserve crew for all possible combinations of simultaneously absent crew whilst also computing their associated probabilities; 4) An additional contribution comes in the form of a model refinement which enables our approach to explicitly take account of the variance present in the to-

tal number of crew absence that can occur in a given scheduling horizon. This refinement improves the accuracy of the descriptive model and as a consequence the quality of the reserve crew schedules that are derived from it; 5) The proposed approach is shown to be a suitable replacement for time consuming simulation approaches whilst also offering the potential for real-time evaluations of alternative recovery actions.

The remainder of the paper is structured as follows: Section 2 reviews the related literature. Section 3 describes the problem which the airlines have to solve. Section 4 presents the probabilistic model which has been developed for this work, and details the important enhancements to this model over that in Bayliss et al. (2012). Section 5 gives experimental results, showing first the improvements in the accuracy of the predictions of the descriptive model, then the benefits that are gained for the decision makers from using this model within a system to determine when to schedule reserve crew. Section 6 concludes with a summary of the main findings and considers future research directions based on this work.

2 Related Literature

Previous literature on reserve crew scheduling includes Gaballa (1979), Dillon and Kontogiorgis (1999) Paelinck (2001) and Sohoni et al. (2006). Gaballa (1979) uses the probabilities of callouts as a guide to reserve sizing. Gaballa (1979) assumes that reserve crew are used when flights are delayed such that the scheduled crew would exceed their maximum duty length if they start the delayed flight. The author observed that the reserve policy used by Qantas at the time meant that overnight delays due to reserve crew unavailability had a 1 in every 166 years chance of occurring, and this was what convinced Qantas that their reserve policy was over-conservative. The presented alternative approach was estimated to save \$600,000 (1979) a year. In contrast to the work of Gaballa (1979) who consider overnight delay probabilities and reserve crew with fixed duty times, this proposed approach relaxes the fixed duty start time constraint and focuses on the minimisation of cancellations due to crew absence.

Dillon and Kontogiorgis (1999) present an approach for pilot reserve crew scheduling that generates reserve duty patterns (reserve pairings) which are then allocated using preferential bidding. They focus on quality of life considerations such as regularity. Dillon and Kontogiorgis (1999) generate call out day pairings of varying lengths which also exhibit regularity. Longer call out pairings allow reserve crew to be used for more different types of pairings including long haul pairings. Generating varying length pairings allows for reserve crew who have different amounts of time off in a given month. In contrast, this work uses reserve pairings which have a fixed length, since this reflects the practices used by the airline on whose data this work is based.

Paelinck (2001) describes a practical approach which was implemented at KLM to optimise cabin crew reserve duties. The approach calculates daily demands for reserves and the expected number of reserve crew remaining each day, and uses a reserve block stacking approach. The aim is to always have standby reserve crew (reserve crew stationed at the airport ready to cover for disrupted crew) available. The work of Paelinck (2001) highlights some of the difficulties associated with the planning and scheduling of reserve crew, including how many should be scheduled, and when and what the best way is to use them in response to disruptions. The work of Paelinck (2001) provided the starting point for Bayliss et al. (2012).

Sohoni et al. (2006) minimise the requirement for reserve cockpit crew (the most expensive crew type) by better predicting the requirement for reserves. They report that reserve crew utilisation is about 40% and claim that, if recurring training is taken into account during crew scheduling, the estimated requirement for reserves will be more accurate because conflicts with recurring training is a leading cause for needing reserve crew.

The problem that is tackled in this work is stochastic in nature so we now consider the various stochastic methods that have been used for similar problems in the past. In particular, there have been a number of stochastic approaches to airline crew scheduling, including Schaefer et al. (2005); Yen and Birge (2006); Weide et al. (2010); Burke et al. (2010); Duck et al. (2012).

Schaefer et al. (2005) solve the traditional crew pairing problem as a set partitioning problem, but the operational costs of crew pairings that are used in the traditional objective function for this problem are replaced with operational costs that are derived from an airline simulator (SimAir). Their approach requires the assumption that the operational cost of one crew pairing can be evaluated in isolation from other crew pairings. They require this assumption because the evaluation of crew pairings using a simulation model can become prohibitively slow. This problem is avoided in this current work by using a probabilistic model to replace the simulation as a method of solution evaluation.

Yen and Birge (2006) describe an iterative approach to increasing crew schedule robustness. In each iteration the traditional crew pairing model is solved, followed by a recourse problem that allows or disallows crew connections that may lead to delay in the next iteration. Pairings that involve aircraft changes hold more potential for delay and are eliminated and not branched upon in their “flight pair branching algorithm”. The recourse problem which calculates “switch delays” does so by considering a set of random scenarios and solving a set of decision variables to compute how delay propagates through the schedule given the most recent solution to the crew pairing problem. Their experimental results indicate that their approach leads to solutions with crew schedules involving fewer aircraft changes and as a result reduces switch delays.

Weide et al. (2010) also consider increasing crew schedule robustness, introducing an iterative approach to solving the integrated crew pairing and aircraft routing problems with the objective of minimising the overall cost and minimising what are termed “restricted aircraft changes”. They take the approach that airline schedule robustness can be increased by decreasing the level of dependency between the crew and aircraft layers of the schedule and therefore reduce the risk of delays caused by delayed crew who have to change to (and delay) another aircraft mid-duty. The probabilistic model that is used for reserve crew scheduling introduced in this current work is complementary to such approaches, since it takes as input the airline schedule including the crew schedule, so will maintain any additional schedule robustness that an approach such as that used in Weide et al. (2010) introduced during other stages of airline schedule planning. The proposed approach then further increases the robustness of the airline’s crewing operations through the process of reserve crew scheduling with the objective of minimising cancellations and delays.

Burke et al. (2010) consider a memetic approach to improving the robustness of airline schedules by allowing flight retimings and aircraft rerouting. They use reliability and flexibility objectives. Reliability measures the ability of the schedule to absorb small disruptions whilst the flexibility objective measures the number swap recovery actions that are possible. Their simu-

lation study of their results showed that the reliability objective had the most impact on schedule robustness whilst the flexibility objective showed some potential for increasing schedule robustness.

Duck et al. (2012) introduce an integrated approach to the crew pairing and aircraft routing problems. Their objective includes cost terms from the usual crew pairing and aircraft routing formulations, plus an expected propagated delay term. To avoid a non-linear stochastic recourse function, the problem is decomposed into separate crew and aircraft routing problems, each with their own recourse problem. Delays for a given schedule are calculated by considering a stochastically generated set of scenarios, where each scenario specifies realised departure and arrival event times. An iterative approach (based on the iterative algorithm of Weide et al. (2010)) is used in which crew pairing and aircraft routing subproblems are solved in each iteration.

The proposed model is also based on a hub-and-spoke network. For further information about this type of network, as well as a description of airline networks in general and a comprehensive treatment of airline operations and delay management, see Wu (2010). For a more general discussion of airline networks and the generation of non-commercial demand data see Akartunal et al. (2013).

As previously mentioned, the predictive model introduced in this current work is used for off-line scheduling, but could also be used in an on-line context to help decision makers decide when to use reserve crew, called the airline schedule recovery problem. Previous work which focusses on the airline schedule recovery problem includes that of Abdelghany et al. (2004), Chang (2012) and Peterson et al. (2012).

Abdelghany et al. (2004) introduce an approach for solving the hub-and-spoke network crew recovery problem which considers crew swaps, reserve crew and deadheading as possible recovery actions. Their Mixed Integer Programming model takes the current crew schedule and disruptions as inputs. They reason that crew disruptions that occur at spokes are often difficult to deal with as there are few connecting flights to spokes, and reserve crew tend to be stationed at the hub. They take the approach of implicitly solving crew disruptions at spokes by solving them at the hub before they occur. This is similar to the model considered in the approach proposed in this paper. Their objective is to recover as many disrupted crew pairings as possible, with the least incurred recovery cost. Their objective function has cost contributions for crew swaps, reserve crew, deadheading and cancellations. Their solution approach solves crew disruptions sequentially in chronological order (earliest disruptions first). This means that the recovery decisions for disruptions are determined as disruptions occur. Such an approach is convenient for airlines who operate a rule-based approach to recovery as opposed to an optimisation based approach. The work of Abdelghany et al. (2004) can be considered to be complementary to this current work, in that this work attempts to assess the robustness of schedules across potential disruptions, whereas Abdelghany et al. (2004) attempts to recover from specific disruptions.

Chang (2012) focusses on the pilot recovery problem. A genetic algorithm approach is presented which takes the original infeasible schedule as input. Crew feasibility constraints in Chang (2012) require a maximum of 10 flying hours per day and 32 flying hours per week. The author introduces an object oriented matrix chromosome structure, where each row corresponds to a pilot and specifies the flights assigned to that pilot. Although Chang (2012) does not directly compare the proposed method with any alternative methods it does appear to be capable of solving large problem instances.

Peterson et al. (2012) address the full airline recovery problem from schedule recovery to passenger re-routing. They argue that sequential approaches to the airline recovery problem naturally lead to sub-optimal solutions because of the conflicting objectives that exist between each problem. On the other hand a fully integrated approach is intractable, a problem which in this case is dealt with by considering a subset of the full airline recovery problem at a time, i.e. only a selection of the disrupted flights and the affected aircraft, crew and passengers are rescheduled at a time.

The method proposed here features a predictive model based on propagating-probabilities. We believe that this approach is more appropriate than some alternative methods of prediction, which we will consider here, because it is designed specifically to reflect the logic of the process that is being modelled. Specifically the predicted cancellation probabilities of flights are functions of the probabilities of cancellations of previous flights and the probabilities that reserve crew remain available given their probabilities of not having been used to cover prior disruptions. The proposed approach uses the structure of the problem to propagate probabilities through a schedule. If the calculations were to be written as a single equation, the result would be a highly non-linear binary decision variable problem. This non-linearity causes problems for many approaches, as well as for solution methods to the sub-problems, as will be explained in Section 4.2. In addition, approaches such as regression methods or neural networks (e.g. Amita et al. (2015)) are unnecessary because the underlying structure of the process is already well understood, so it is not necessary to fit the parameters of a general model to data. However a possible future study might consider a neural network approach that uses the proposed model to generate training data. The neural network could then facilitate fast approximate reserve crew schedule evaluation. In addition to the problems caused by the non-linearity of the problem, prediction methods such as SVMs (e.g. Gu et al. (2015); Gu and Sheng (2017)) are not appropriate because this problem is concerned with probability prediction rather than binary classification.

In this work the proposed model for evaluating candidate reserve crew schedules is utilised within heuristic and metaheuristic algorithms including simulated annealing Kirkpatrick et al. (1983) and genetic algorithms Goldberg (1989). Other work is concerned with the development and testing of more advanced metaheuristic algorithms such as that introduced in Deng et al. (2012) which features a hybridised metaheuristic based on ideas from genetic algorithms, particle swarm optimisation and ant colony optimisation. Other novel developments in metaheuristic algorithms include Deng et al. (2015, 2017b,a); Xue et al. (2018).

In previous work which considered the problem of airline reserve crew scheduling under uncertainty, a probabilistic model of crew absence and the reserve crew who are used to cover absent crew was introduced Bayliss et al. (2012). In this current work this model is extended to remove a variety of simplifying assumptions. The previous model is detailed in Section 4.2, the improvements since Bayliss et al. (2012) are explained in Section 4.3, then the remainder of Section 4 details the improved model.

In other previous work, Bayliss et al. (2013) introduce a probabilistic model of propagating crew-related delay (but not crew absence) and the effect that reserve crew have on absorbing crew-related delays and any resultant knock on delays. A matrix was used to model the propagation of crew-related delay, with columns corresponding to the source flights of the crew-related delay and rows corresponding to the flights which are affected by the delay. In the probabilistic crew-related delay model in

Bayliss et al. (2013), use of reserve crew meant teams of reserve crew completely replacing a team of crew who were on a delayed connecting flight. The model proposed in this current paper provides for much more flexibility.

An alternative approach to the probabilistic models of crew-related disruptions and reserve crew usage was investigated in Bayliss et al. (2017), which introduced a mixed integer programming approach to reserve crew scheduling (*MIPSSM*, mixed integer programming simulation scenario model). In Bayliss et al. (2017) disruption scenarios were derived from an airline simulation, which were then used to form the objective and constraints of a mixed integer programming formulation with the goal of finding the reserve crew schedule that minimises delay and cancellations over all of the disruption scenarios included in the formulation. In contrast to Bayliss et al. (2012) and Bayliss et al. (2013), the *MIPSSM* in Bayliss et al. (2017) also accounts for both absence and delay disruptions as this paper does. The *MIPSSM* approach from Bayliss et al. (2017) is compared with the models introduced in this work in Section 5 further indicating the benefits of the approach described in this paper. We refer the reader to Bayliss et al. (2017) for a review of additional literature in robust optimisation.

3 Reserve Crew Scheduling Problem Description

The structure of the problem being tackled by airlines is described in this section.

Given the inputs: a scheduling horizon, an input airline schedule (crew and aircraft routings), probabilities of crew absence and an assumed on-line reserve policy, the objective is to assign standby duty periods to reserve crew in a pre-emptive fashion (off-line) to minimise the total expected number of flight cancellations and reserve crew induced delays which occur during the scheduling horizon. The constraints of this problem include that reserve crew are on standby duty for a fixed amount of time after their duty periods begin and can only be used to replace disrupted crew if the expected final arrival time of the flight duty they are used for finishes within their assigned duty length plus any additional duty length slack (permissible overtime). Given this overview of the problem, we now provide the important definitions and describe the key problem aspects and assumptions that are taken into account in our solution approach.

In the airline industry, flight legs are single flights from one location (origin) to another (destination). Crew schedules consist of sequences of flight legs. A sequence of flight legs to be completed within a single day is referred to as a duty. A sequence of duties which begins and ends at the crew's home base is referred to as a crew pairing. The required crew for a flight include pilots and cabin attendants. Different aircraft types (called fleets) each have specific minimum crew requirements. This work focusses on a single crew type and assumes a single fleet type (Section 6 considers the more general case, but scheduling fleet types independently is actually common in practice). Therefore, the crew requirements for all flights are equal. The example case of 4 crew per flight is considered in this work, but the model holds for different numbers of crew as well. Additionally, crew serve out their crew pairings in teams of crew who stay together throughout their crew pairings.

This work considers the situation where reserve crew are on standby duty at the hub station of an airline which operates a single hub and spoke network. In a hub and spoke network all flights involve the hub station as an origin or a destination. This is

efficient from an airline's perspective as it maximises the number of connections that can be made, but passengers often have to change aircraft at the hub station to get to different destinations.

Each reserve crew has a start time for their standby duty, and a duty length. This work focusses on the problem of allocating the standby duty periods for reserve crew who are used on the day of operation to replace absent crew so that flight cancellation is avoided. The key contribution of this work is an improved model for crew absence uncertainty, which allows reserve crew schedules to be evaluated. Two assumptions are made within this model: 1) Absent crew miss their entire assigned crew pairing. If absent crew then become available they are often assigned to other open pairings which have no crew assigned to them; 2) All crew have a probability of being absent. In this work a constant 1% chance of crew absence for each individual member of crew is assumed. A value of 1% was used since the value in the actual crew absence data was approximately this, however the model could equally well utilise other values, and is general enough to cope with separate probabilities of crew absence for each individual crew, if desired.

A metaheuristic approach has been developed to use this evaluation model to find reserve crew schedules which are predicted to perform well. This work considers the case where the number of reserve crew available for scheduling is fixed, and the possible start times for reserve crew standby duties are discretised according to the scheduled departure times at the hub. Reserve crew are feasible for disrupted duties which can be completed by the reserve crew within their maximum allowable duty length, which includes permissible overtime.

Flights which are delayed beyond a cancellation threshold (CT) of 180 minutes are assumed to be cancelled. The same assumption is used in Rosenberger et al. (2003). Thus, reserve crew feasibility also requires that reserve crew begin their standby duty period no later than 180 minutes after the scheduled departure time of the affected flight. If a flight is cancelled, the crew absence rolls over to the next scheduled hub departure on the same crew pairing, at which time there is another opportunity to replace the absent crew with reserve crew to avoid further flight cancellations.

All of the absent crew affecting a given departure have to be replaced in order to avoid a flight cancellation. Reserve crew are only used if all of the absent crew can be replaced. It is assumed that crew absence becomes known at the scheduled departure time of the first flight in a crew pairing, and reserve crew are considered for use in earliest start time order.

The objective here is to schedule reserve crew in such a way that the expected number of cancellations due to crew absence is minimised. For the purpose of this evaluation, delays are counted as partial cancellations, so that they are avoided. Reserve induced delays occur when reserve crew have standby duty start times that are greater than the departure times of the departures they are used for, so that the flight is delayed until the reserve is available. A term is included in the objective function to penalise these reserve induced delays. A delay threshold (DT) of 15 minutes is used (also used by Sohoni et al. (2011)), which means that the delay stopwatch is started after 15 minutes and delay below this is not considered delay.

The proposed model is a predictive model of cancellations due to crew absence, using as input the probabilities that the assigned crew are absent and calculating the reduction in cancellations due to crew absence for any given reserve crew schedule. The model is used within an off-line heuristic search for good schedules within this research, solving the problem that the airlines face of determining when to schedule the reserve crew. (The quality

of the results is here evaluated using a Monte-Carlo simulation, with the same constraints and objective function.)

In addition to solving the off-line planning problem, airlines also often have to consider an on-line problem of deciding on the day of operation when to actually use the available reserves. The developed model would also be valuable for this purpose, enabling the decision makers to see the effects of different decisions very quickly, without the need for lengthy simulation. Despite being under consideration at the time of writing, the application of the model for this problem is not evaluated here, however we note that in an on-line application the primary difference is that the probability matrix P of crew absences (see Section 4.2) would be updated as crew absence outcomes are realised. i.e. realised outcomes have a probability of 1. The model presented here is therefore of use for other variants of the reserve crew scheduling problem beyond the off-line problem for which it is evaluated in this paper.

4 Probabilistic Model

In contrast to the work of Schaefer et al. (2005) and that of Yen and Birge (2006), which capture uncertainty in the form of randomly generated scenarios and then evaluate how the solutions perform in each of those scenarios, the probabilistic approach proposed in this work evaluates solutions in a single probabilistic scenario in which all disruptions have a probability of occurring. This probabilistic model is explained in this section.

Due to the number of definitions needed by the model, these have been collected into Section 4.1. The similar (simplified) model which was presented in Bayliss et al. (2012) is then summarised in Section 4.2. Following this, the various improvements to the model are detailed in Section 4.3 and the model changes to support these improvements are explained in Sections 4.4 and following.

The aim of this work is to provide a model which is as effective as a simulation/scenario-based approach without requiring the same amount of time as such an approach. The prediction accuracy of the model will be evaluated in Section 5.3 and Section 5.4 considers the use of the model within a reserve crew scheduler to solve the reserve crew scheduling problem.

4.1 Definitions

The variable and constant definitions are summarised in the following table, along with details of which sections they are used in:

$a_{d,e}$:	The probability that a total of e reserve crew are available to cover a total of e absent crew affecting departure d (Section 4.4)
A_d	:	Scheduled arrival time of departure d (Section 4.5)
CM	:	A vector of cancellation measures for reserve induced delay (Section 4.6)
CM_d	:	An element of CM , cancellation measure of reserve crew use induced delay at departure d (Section 4.6)
CT	:	Cancellation threshold (maximum delay before a flight is cancelled). A value of 3 hours was used for the experiments in this paper. (Section 4.6)
D_d	:	Scheduled departure time of departure d (Section 4.5)
$delExp$:	Delay exponent used in the delay cancellation measure function (a value of 2 is used in this work) (Section 4.6)
DL	:	Maximum duration of a reserve standby duty period (Section 4.5)
EDT_d	:	Expected departure time for departure d (Section 4.6)
F_d	:	Crew pairing assigned to hub departure d (Section 4.7)
$Feas_{t,d}$:	Binary matrix for the feasibility of a reserve crew member with start time index t covering crew absence affecting departure d (Section 4.5)
$L_{l,d}$:	departure number of the last flight of the day of crew pairing l on the day of departure d (Section 4.5)
$leafNodes$:	The number of leaf nodes currently in the reserve crew combination tree (Section 4.8)
$Leaves$:	The set of leaf nodes in the reserve crew combination tree at any given time (Section 4.8)

M	:	The set of reserve crew feasible for a given departure to cover crew absence (Section 4.7)
$maxCA$:	Maximum number of crew that can be absent from a pairing. Equals the number of scheduled crew in each crew team (Section 4.8)
n	:	Number of hub departures in the airline schedule (Section 4.7)
N_{δ}^{res}	:	δ^{th} reserve crew node (Section 4.8)
$N_{\delta}^{\text{res}, \text{gen}}$:	Number of reserve crew in the reserve crew combination corresponding to node δ in the reserve combination tree (Section 4.8)
$N_{\delta}^{\text{res}, \text{par}}$:	Parent node of node δ in the reserve crew combination tree (Section 4.8)
$N_{\delta}^{\text{res}, \text{res}}$:	Reserve number of the reserve crew member corresponding to node δ in the reserve crew combination tree (Section 4.8)
$nodeProb$:	Probability that a given combination of reserve crew are simultaneously available for covering crew absence (Section 4.8)
P	:	The matrix of probabilities that a given number of crew will be missing from a given flight (Section 4.4)
$p_{d,e}$:	An element of P , the probability that e crew are unavailable for departure d (Section 4.4)
Q	:	The matrix of probabilities that a given number of crew will be missing from a given flight before reserve crew are used (Section 4.4)
$q_{d,e}$:	An element of Q , the initial probability of e absent crew at departure d before the affects of a reserve crew schedule are taken into account (Section 4.4)
$r_{d,k}$:	Probability that reserve crew member k is available to cover crew absence affecting departure d (Section 4.4)
R	:	Number of reserve crew in a reserve crew schedule (Section 4.4)
$ResCom$:	Vector containing the combination of reserve crew corresponding to a given node in the reserve crew combination tree (Section 4.6)
$u_{d,k}$:	Probability that reserve crew member k is used to cover crew absence affecting departure d (Section 4.4)
X	:	Reserve crew schedule (Section 4.2)
X_k	:	Start time index of the k^{th} reserve scheduled to begin a reserve pairing (Section 4.2)
ξ	:	Number of nodes currently in the reserve crew combination tree (Section 4.8)

4.2 Basic Probabilistic Model

We first summarise the simplified probabilistic crew absence model (*SPCAM*) from Bayliss et al. (2012). The aim of the model is to determine the probability of cancellation due to crew absence (p_d) for each flight (d). Given a reserve crew schedule, the model works through the flights in the schedule, considering the probabilities of crew absence and which reserve crew are available, and using reserve crew where useful. Each reserve crew member (k) is modelled as having a probability ($r_{d+1,k}$) of remaining available immediately after the possibility of being used to replace absent crew affecting each flight d . Reserve crew are assumed to be feasible to replace absent crew for the first L flights that departed during their standby duty period. x_k denotes the standby duty start time index of reserve crew k , where potential start times are discretised according to the airline's scheduled departure times.

Based on this, Algorithm 1 calculates the effect that any given reserve crew schedule would have on reducing the probabilities of cancellations due to crew absence. i.e. it calculates the probabilities of cancellations due to crew unavailability (p_d)—regular or reserve crew.

Algorithm 1 Scheme for calculating crew unavailability probabilities given a reserve crew schedule

```

1:  $P = Q$ 
2: for  $k = 1$  to  $R$  do
3:    $r_{x_k,k} := 1$ 
4:   for  $d = x_k$  to  $\min(n, x_k + L - 1)$  do
5:      $r_{d+1,k} := r_{d,k}(1 - p_d)$ 
6:      $p_d := p_d(1 - r_{d,k})$ 
7:   end for
8: end for

```

Line 1 of Algorithm 1 initialises the probabilities of cancellation due to crew unavailability (P) to the probabilities that the crew assigned to each flight are absent (the probability vector Q). In line 2 the algorithm considers each reserve crew member in turn, in earliest start time order (X is sorted in non-decreasing order of start time index).

Each reserve crew's probability of availability is initially set to 1 (line 3). For each reserve crew the algorithm cycles through each flight for which that reserve could feasibly be used to replace absent crew affecting that flight (line 4). For each such flight, the reserve crew's probability of remaining available for subsequent flights is calculated to allow for the possibility that they are required to replace absent crew affecting the flight. This is their probability of still being available for flight d , multiplied by the probability that the reserve crew is not required to replace absent crew affecting flight d (line 5). At the same time the probability that flight d is cancelled due to crew unavailability is updated to allow for the possibility that the reserve is available to replace absent crew affecting flight d (line 6).

The order in which reserve crew are considered (line 2) has the effect of modelling a preference order for the use of reserve crew, which in this case is assumed to be the earliest start time order. This rule of thumb reserve order policy maximises the total remaining reserve crew standby duty time. Lines 5 and 6 represent the fundamental equations of the *SPCAM*.

A heuristic approach will be used later to solve this problem. Before considering (in Section 4.3) the model improvements which have been made, we first consider here the difficulties that exact approaches such as (mixed integer) linear programming and dynamic programming would face, and hence why heuristic solution methodologies are appropriate for solving the (extended) version of Algorithm 1 to find disruption minimising reserve crew schedules. If the reserve crew schedule is encoded as a binary vector, and Algorithm 1 is written as a closed form equation, the resulting mathematical programming formulation is a highly non-linear function of the binary reserve crew schedule decision variables. In fact, there would effectively be one non-linear term for each possible reserve crew schedule. This makes the problem impractical for solution by linear programming methods. An alternative approach may be a dynamic programming formulation of the problem. We note, however, that with a state definition based on the number of reserve crew remaining to be scheduled and current departure number, the Markov property would not hold, since both future disruptions and the availability of previously scheduled reserve crew are strongly dependent upon the exact duty start times allocated to previously scheduled reserve crew. This interaction between previous and later decisions ensures that the state space for a dynamic program will be impractically large¹, resulting in an intractably large problem to solve. We will therefore opt for meta-heuristic solution techniques in Section 5.

4.3 Improvements Over Previous Model

The assumptions from Bayliss et al. (2012) that are addressed in this work are as follows:

1) **Crew absence was accurately modelled with a single probability per departure.** This simplification did not allow for the possibility that a number of the crew assigned to a flight can be absent simultaneously. In this improved model the single probability of crew absence for each flight is replaced with a probability distribution containing the probabilities that different numbers of crew are simultaneously absent. As a result of considering the possibility of multiple crew from a crew team being simultaneously absent, reserve crew feasibility depends on the combined feasibility of a group of individual reserve crew. Furthermore, flight cancellation due to crew unavailability can

¹The state space would consist of all possible subsets of all possible reserve crew schedules with start times bounded by each possible reserve duty start time interval plus the cancellation threshold

only be avoided if all absent crew are replaced, one for one, with reserve crew.

2) **The probabilities of crew absence affecting flights in a schedule were independent.** This simplification ignores the structure of an airline’s schedule, that is, crew teams are often assigned to multiple flights, and therefore the probabilities of crew absence affecting those flights are dependent upon one another. In this work the assumption is made that crew, if absent, are unavailable for all flights in their assigned crew pairing. The improved probabilistic crew absence model therefore better takes the structure of the crew schedule into account.

3) **Reserve crew were feasible to cover for crew absence disruptions affecting a fixed number of departures after the beginning of their standby duty.** The simplified probabilistic crew absence model ignored the details of the airline’s schedule, such as departure times and arrival times. In this work, reserve crew feasibility depends on whether the disrupted duty is expected to finish within the reserve crew standby duty period.

4) **Reserve crew were not feasible to cover crew absence disruptions affecting flights whose scheduled departure time was before the start of their standby duty.** This simplification did not allow for the possibility that a flight which was affected by crew absence could wait for reserve crew to begin their standby duties. The relaxation of this simplification introduces the possibility that using reserve crew to avoid flight cancellation can introduce a delay. In this work reserve use induced delays are incorporated into the objective function using a function which converts delays into a measure of cancellations. The delay cancellation measure function is introduced in Section 4.6.

4.4 Enhanced Probabilistic Model

The *SPCAM* can be extended to the case where we do not assume that crews come in inseparable teams. To allow for this, the improved probabilistic crew absence model (*CAM*) requires as input a discrete probability distribution which specifies the probabilities that different numbers of crew are simultaneously absent for each departure.

Let Q now be a matrix containing the set of all crew absence distributions, so that $q_{d,e}$ corresponds to the probability that a total of e crew are simultaneously absent for a departure d , whereas $p_{d,e}$ is the probability of crew unavailability—assigned crew or reserve crew.

The required crew absence distributions for each team of crew are calculated using binomial distribution probabilities, based on the assumed 1% chance of any given member of crew being absent and the assumption that crew teams consist of 4 individuals (see Section 3). For example, the probability that exactly two crew are simultaneously absent ($5.881E-4$) is calculated as $6 \times ((0.99^2) \times (0.01^2))$, since there are 6 ways that exactly two out of four crew can be absent simultaneously—6 is a binomial coefficient. These probabilities turn out to be (0.9606, 0.0388, 5.881E-4, 3.96E-6, 1E-8) respectively for (0,1,2,3,4) crew being simultaneously absent.

In the *CAM* the fundamental equations, equivalent to those of the *SPCAM* given in lines 5 and 6 of Algorithm 1, are replaced with the following.

$$\begin{aligned} r_k^{d+1} &:= r_k^d - u_{d,k} & (1) \\ p_{d,e} &:= p_{d,e} (1 - a_{d,e}) & (2) \end{aligned}$$

Where $a_{d,e}$ is the probability that a total of e reserve crew are simultaneously available at departure d .

Equation 1 gives the probability that reserve k remains available for subsequent use given that they have a probability of $u_{d,k}$

of being used to cover absence at departure d . Equation 2 states that the probability that a total of e crew are unavailable for departure d depends on the probability that a total of e crew are absent in the first place and the probability that a total of e reserve crew are not available (simultaneously) to cover the absence affecting departure d . Just as in Algorithm 1, Equations 1 and 2 are applied in an iterative fashion to probabilistically model reserve crew use (in a propagating probabilities approach) see Algorithm 2. The calculation of $a_{d,e}$ requires the enumeration of all combinations of e reserve crew which are feasible simultaneously for departure d , which is considered in detail in Section 4.8. Just as in the *SPCAM*, the *CAM* requires a preference order for specifying the order in which reserve crew are to be considered for use for covering crew absence disruptions affecting each departure. This approach captures the feature that the probability that a lower preference combination of reserve crew are used to cover crew absence depends on the probability that higher preference reserve crew combinations are not available to cover the same disruption. Algorithm 2 (which is the focus of Section 4.7) is to the *CAM* what Algorithm 1 is to the *SPCAM*.

4.5 Reserve Crew Feasibility

This section defines the feasibility of reserve crew scheduled at different times for covering flights which may be affected by crew absence, and is required to replace the simplified model for reserve crew feasibility used by the *SPCAM* (see line 4 of Algorithm 1). In contrast to the *SPCAM*, reserve feasibility in the *CAM* depends on the exact structure of crew pairings and the duration of the cancellation threshold (CT).

$$Feas_{t,d} = \begin{cases} 1 & D_t < D_d + CT \text{ and } D_t + DL \geq A_s \\ & (s = L_{F_d,d}) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Equation 3 states that reserve crew with start time index t are feasible for a disrupted departure d , if: 1) they begin their standby duty before the cancellation threshold of the disrupted departure, and 2) their duty finishes ($D_t + DL$) at or after the final arrival time of the disrupted pairing (A_s). In the event that e crew are absent, a combination of e reserve crew is feasible if and only if each of the individual reserve crew are feasible. The feasibility of a reserve crew with start time index t being available to cover crew absence disrupted departure d can be pre-calculated and stored in the form of a binary matrix.

4.6 Reserve Use Induced Delay

The *SPCAM* did not allow for the possibility of reserve crew being used for crew absence disruptions which occur before the start of their standby duties. The *CAM* relaxes this constraint, which means that delays can be introduced when using reserve crew who begin their standby duties after the scheduled departure time of a crew absence disrupted departure. This section adds a term to the objective function of the *CAM* that penalises reserve induced delays.

This model is referred to as the static delay model (*SDM*), which is simply the *CAM* with a delay penalty term in the objective function (see Equation 9 of Section 4.9 for the objective function based on this model). To penalise reserve induced delay (Equation 4) a delay cancellation measure function is used (Equation 5). This approach means that the *SDM* returns a single objective value for any given reserve crew schedule, i.e. expected cancellations. EDT_d is the expected delay of departure d before the effects of reserve crew are considered, which accounts for

delays due to unexpectedly high journey times of a prior flight, which cannot be prevented through reserve crew use. Such delays can be estimated from a simulation, in which no reserve crew are scheduled. The use of expected delays EDT_d in Equation 4 motivates the name: *static* delay model. In contrast to a static model, a dynamic model of delay would mean that EDT_d is responsive to the given reserve crew schedule, because reserve induced delays can have knock on effects in terms of how delays propagate and how airlines respond to future delays, and is an interesting topic for further research.

Equation 4 gives the delay associated with using a given combination of reserve crew ($ResCom$) to cover crew absence affecting departure d . The delay depends on the utilised reserve crew (k) with the latest duty start time.

$$delay = \max_{k \in ResCom} (D_k - D_d, EDT_d) \quad (4)$$

$$Cancellation\ measure = \left(\frac{delay}{CT} \right)^{delExp} \quad (5)$$

Equation 5 gives the cancellation measure of a delay as the ratio of the *delay* relative to the cancellation threshold (CT), raised to the power *delExp* (delay exponent). The same delay cancellation measure function was used in Bayliss et al. (2017). The choice of delay exponent (*delExp*) represents the subjective equivalence of the perceived level of disruption between delays of different sizes and a flight cancellation. The cancellation measure of a delay is 1 if the delay equals the CT . The higher the value of *delExp* the less delays are penalised. As *delExp* approaches infinity the delay cancellation measure approaches a step function, returning 1 if a flight is cancelled due to delay, and 0 otherwise. The effect of the value of *delExp* which is used in a reserve crew scheduling application is shown in Section 5.1.

$$CM_d := CM_d + g \left(\frac{delay}{CT} \right)^{delExp} \quad (6)$$

Equation 6 gives the objective value contribution (penalty) associated with reserve induced delay due to a combination of reserve crew who have a probability of g of being used to cover for absent crew affecting departure d . The calculation of g is addressed in section 4.8.

4.7 Evaluating Expected Cancellations Associated With A Given Reserve Crew Schedule

Algorithm 2 outlines the procedure followed by the *CAM* when evaluating the expected number of cancellations due to crew absence associated with a given reserve crew schedule (X). Algorithm 2 is analogous to Algorithm 1, Algorithm 2 allows for the simplifications that were made in Algorithm 1. In general, Algorithm 2 considers each scheduled departure in order. For each, it enumerates feasible combinations of reserve crew and their associated probabilities of being considered for use. The probabilities that different numbers of reserve crew are simultaneously available are used to update the probabilities that flights are cancelled due to crew unavailability. The probabilities that reserve crew remain available for subsequent disruptions depend on the probabilities that they are used for crew absence affecting the given departure.

In more detail, the algorithm first initialises (line 3) the probabilities of different numbers of crew being unavailable for each departure (P) to the probabilities of different numbers of crew being absent for each departure (Q). Then (line 4) the reserve crew availability probabilities are initialised to 1. The algorithm

Algorithm 2 Outline of reserve crew schedule evaluation procedure

- 1: **Inputs:** airline schedule, the assumed reserve policy, crew absence probabilities (Q), expected delays before reserve recovery
 - 2: **Outputs:** For all flights: cancellation probabilities, reserve use induced delay cancellation measure contributions
 - 3: $P := Q$
 - 4: $r_k^{xk} := 1, \forall k \in \{1 \dots R\}$
 - 5: **for** $d = 1$ to n **do**
 - 6: Reset a and u
 - 7: $M = \{\text{Feasible reserves for departure } d \text{ in earliest start time order}\}$
 - 8: For all feasible reserves in M generate all reserve combinations containing between 1 to $|p_d|$ reserve crew
 - 9: Determine the probability that each combination is used, given that reserves are used in earliest start time order
 - 10: Determine (a) the total probability of different numbers of reserve crew being simultaneously available
 - 11: Determine (u) the probability that each individual reserve is used to cover crew absence in departure d
 - 12: Update probabilities of cancellation due to different numbers of crew absence and reserve availability for subsequent crew absence
 - 13: $p_{d,e} := p_{d,e} (1 - a_{d,e}), \forall e \in \{1 \dots |p_d|\}$
 - 14: $p_{w,e} := p_{d,e}, \forall w \in \{\text{subsequent departures assigned to crew pairing } F_d\}, \forall e \in \{1 \dots |p_d|\}$
 - 15: $r_{d+1,k} := r_{d,k} - u_{d,k}, \forall k \in M$
 - 16: **end for**
-

then considers each scheduled departure (line 5) in earliest departure time order. For each departure, all combinations of the reserve crew which are feasible to cover absent crew affecting that departure are generated (line 8) and their probabilities of being considered for use, given that more preferable combinations are not available, are calculated (line 9). The probabilities that different numbers of reserve crew are simultaneously available (line 10) are used to calculate the probabilities that crew absence disruptions affecting the given crew pairing still go uncovered (line 13). Line 14 updates the probabilities of crew unavailability for downstream flights on the same crew pairing, i.e. to the probabilities that were calculated on line 13. This step captures the assumption stated in the second part of Section 4.3 regarding the dependency of the crew absence probabilities between flights which are assigned to the same crew team. After this, the probabilities that each individual reserve crew is used to cover absence at the given departure (which are calculated on line 11) are used to update the probabilities that each reserve crew remains available for subsequent crew absence disruptions (line 15).

The details of how the feasible combinations of reserve crew are actually generated (line 8) and their corresponding probabilities calculated (line 9) are the subject of Section 4.8.

4.8 Enumerating Feasible Combinations Of Reserve Crew And Associated Probabilities

Lines 8 to 11 of Algorithm 2 involve enumerating feasible combinations of reserve crew of different sizes and calculating their probabilities of actually being utilised. These probabilities depend on the probabilities that different numbers of crew are not available and the probabilities that more preferable combinations of reserve crew are available for the same disruptions. Reserve combination preference is defined by the reserve use order pol-

icy, which in this case is assumed to be the earliest start time order, so as to minimise reserve induced delay. Note that any other order based policy could be used instead.

Algorithm 3 enumerates the feasible combinations of reserve crew for each possible crew absence disruption. It turns out that simply enumerating all combinations of reserve crew of different sizes also yields combinations that in reality would never occur, given the reserve use policy. Such combinations include:

1. Combinations that have been generated for a previous flight in the same crew pairing, as crew absence is covered at the earliest opportunity since it makes no sense to hold reserve crew when they can be used to cover crew absence to prevent flight cancellation. Such combinations are filtered out by line 15 of Algorithm 3.
2. Combinations involving non-consecutive reserve crew numbers with identical duty start times or identical flight feasibility. This is because reserve crew are used in order (which is the order specified by the given reserve policy) and an ordering is applied to reserve crew with the same start time, to reduce symmetries. Such combinations are not generated by Algorithm 3, see lines 24 to 26.

Another aspect of reserve combination generation that requires careful consideration is the derivation of the probabilities that given combinations are considered for use, given that they pass reserve combination filters 1 and 2 (above). For example, suppose that using reserve crew 1, 2, 3 and 4 are feasible recovery actions when 2 crew are absent from a given crew team, then a feasible reserve combination such as (1,4) (which implies 2 and 3 have different start times/flight feasibility to 1 and 4) has an associated probability of $(r_1(1-r_2)(1-r_3)r_4)$ of being considered for use to cover the disruption. However, the $(1-r_2)$ term can be removed if the reserve crew combination (1,2) was a feasible combination for a previous flight on the same crew pairing, because had this been the case, the reserve crew combination (1,2) would have been used to cover the two absent crew at that time. In this case, the use of reserve 4 does not depend on reserve 2 not being available, in fact their usage is mutually exclusive to each other. The same reasoning applies to the $(1-r_3)$ term. This concept is referred to as *reserve non dependency* later on, and is enforced in lines 17 to 21 of Algorithm 3.

Algorithm 3 outlines the procedure for generating feasible combinations of reserve crew and calculating their associated probabilities of being used. The algorithm is based on building a tree of nodes, where nodes correspond to particular feasible reserve crew, and paths from the root to a leaf correspond to combinations of reserve crew.

Figure 1 illustrates how starting from a root node (top) the reserve crew combination tree is generated in stages, where in each stage the next preferred reserve crew is added to the tree. The node probabilities for the new reserve combinations generated in each iteration are stated below the newly generated nodes. The new reserve crew combinations generated at each stage are listed at the bottom of the diagram.

In Algorithm 3, *Leaves* denotes the set of leaf nodes and *LeafNodes* the number of leaf nodes in the reserve combination tree at any given stage of the algorithm ($|Leaves|$). N_ξ^{len} corresponds to the number of reserves in the combination of reserves beginning at the root node and ending at node ξ . N_ξ^{par} is the parent node of node ξ in the tree. N_ξ^{res} gives the reserve number

Algorithm 3 Generation of reserve combinations and associated probabilities at departure d

```

1: Inputs: departure number ( $d$ ), the probabilities of reserve
   crew availability ( $r$ ), flights in each crew pairing ( $F$ ), reserve
   crew feasibility ( $Feas$ )
2: Outputs: probabilities that different numbers of reserve
   crew are available to replace different numbers of simultane-
   ously absent crew ( $a$ ), the probability that each reserve crew
   member is used for the given departure ( $u$ )
3: Create root node  $N_1$  corresponding to the empty reserve crew
   combination, with no parent node of its own, path length 0
   and node probability 1,  $N_1^{res} := null$ ,  $N_1^{par} := null$ ,  $N_1^{len} := 0$ ,
    $N_1^p := 1$ 
4:  $\xi := 1$  (nodes used)
5:  $Leaves \leftarrow N_1$ 
6:  $leafNodes := |Leaves|$  (nodes to be branched on)
7: for each  $k \in M$  do
8:   for  $\delta = 1$  to  $leafNodes$  do
9:     if  $N_\delta^{len} < maxCA$  then
10:      Branch on  $N_\delta$  with reserve  $k$ 
11:      increment  $\xi$ 
12:       $N_\xi^{res} := k$ ,  $N_\xi^{par} := N_\delta$ ,  $N_\xi^{len} := N_\delta^{len} + 1$ ,  $N_\xi^p := N_\delta^p \times$ 
        $r_k$ 
13:       $Leaves \leftarrow N_\xi$ 
14:       $ResCom =$ The combination of reserve crew corre-
       sponding to the path from  $N_\xi$  to the root node
15:      if  $ResCom$  was not feasible for any previous flight on
       crew pairing  $F_d$  then
16:         $nodeProb := N_\xi^p$ 
17:        for each feasible reserve crew  $s$  which is not in
        $ResCom$ , with start time  $\leq D_{X_k}$  do
18:          if replacing reserve  $k$  with reserve  $s$  in  $ResCom$ 
           results in a combination of reserve crew that was
           feasible for any previous flight on crew pairing
            $F_d$  then
19:             $ResCom$  probability does not depend on
           reserve  $s$  not being available, therefore
            $nodeProb := \frac{nodeProb}{(1-r_s)}$ 
20:          end if
21:        end for
22:        Update the probabilities that  $N_\xi^{len}$  reserve are avail-
           able ( $a_{d,N_\xi^{len}}$ ), that each reserve in  $ResCom$  is used
           for the departure ( $u_{k \in ResCom}$ ) and add a delay can-
           cellation measure contribution corresponding to
            $ResCom$ 
23:        end if
24:        if  $N_\delta^{res}$  has identical feasibility to reserve  $k$  then
25:          Remove  $N_\delta$  from  $Leaves$ 
26:          decrement  $\delta$ 
27:        else
28:           $N_\delta^p := N_\delta^p \times (1-r_k)$ 
29:        end if
30:      end if
31:    end for
32:     $leafNodes := |Leaves|$ 
33: end for

```

corresponding to node ξ . Line 4 defines the root node as node 1, corresponding to a reserve combination of 0 reserves, without it's own parent node and a node probability of 1. Lines 5 and 6 add the root node to the set of nodes (*Leaves*) that are to be branched on (with nodes corresponding to the first feasible reserve) in the first iteration of the algorithm.

The reserve combination tree is grown by branching on each leaf node with branch nodes for each feasible reserve crew in turn, in earliest start time order (lines 7 and 8). This means that each path from the root node to a leaf node defines a combination of reserve crew listed in earliest start time order, with no repeat reserve crew. Additionally, no leaf nodes are more than ‘the maximum number of absent crew’ away from the root node (line 9), as such reserve crew combinations are never required. ξ is the number of nodes in the reserve crew combination tree at any given time. So node ξ always corresponds to the newest reserve crew combination (*ResCom*) generated by the tree.

The probability that the reserve combination (*ResCom*) corresponding to node N_ξ is used depends on the probability that more preferable reserve crew are not available (enforced on line 28) and whether or not the reserve combination is subject to the *reserve non dependency* described above (enforced on lines 17 to 21).

Every time a node is branched on by a new reserve node, the node which was branched on remains a leaf node, but the node probability is updated so that it corresponds to the newest reserve not being available (line 28). The branch node then corresponds to the combinations which that reserve is a member of. The branch node is added to the set *Leaves*, to be branched on by subsequent reserve crew. Given the probability (*nodeProb*) that the reserve crew combination *ResCom* is considered for use, line 22 updates: the probabilities that the individual reserve crew in *ResCom* are used (*u*) to cover crew absence in the given departure (*d*); the probability that a total of N_ξ^{len} reserve crew are available (*a*) to cover N_ξ^{len} absent crew at departure *d*; and the delay cancellation measure contribution for departure *d*, corresponding to *ResCom* (see Equations 4 and 6). Lines 24 to 29 ensure that reserve combinations that fall into the category of reserve combination filter 2 (see above) are not generated. Line 32 sets the number of nodes that are to be branched on when nodes corresponding to the reserve with the next highest start time are added to the reserve crew combination tree.

In this work metaheuristic algorithms (Section 5.2) use Algorithm 3 to evaluate candidate solutions in the search for a cancellation and delay minimising reserve crew schedule. The algorithmic complexity of such algorithms depends on the number of times Algorithm 3 is executed, which is equal to *n* (the number of scheduled departures from the hub station) multiplied by the evaluation budget *E* of the metaheuristic algorithm in question. The algorithmic complexity is closely approximated and bounded above by the following:

$$O\left(E \sum_{i=1}^n \sum_{j=1}^{|M_i|} 2^{j-1}\right). \quad (7)$$

The computational cost of a single implementation of Algorithm 3 depends on the size of the reserve crew combination tree (Figure 1) which depends on the number of reserve crew whose duty period overlaps with each hub departure *i* which is denoted $|M_i|$.

The solution space of this reserve crew scheduling problem is considered in the next section.

4.9 Solution Space

The number of possible reserve schedules is as follows.

$$\sum_{j=\text{ceil}(\frac{R}{\text{MaxCA}})}^R \frac{n!}{j!(n-j)!} y(j, R) \quad (8)$$

Equation 8 gives the number of ways *R* reserve crew can be assigned to the *n* different possible reserve standby duty start time indices (scheduled departure times), where no more than *MaxCA* (maximum number of crew absent from each crew pairing) reserve crew are assigned to any individual start time index, where $y(j, R)$ is the number of combinations of *j* integers ($1 \leq \text{integers} \leq \text{MaxCA}$) that sum to *R*. For the case where no restriction is placed on the number of reserve crew that can begin duties at the same time the $y(j, R)$ values are in the ‘Bell number’ sequence.

The reserve crew schedule *X* specifies the start time index in start time order of each reserve crew scheduled. Where a start time index X_k corresponds to the beginning of a standby reserve pairing, where standby duties begin at time D_{X_k} daily. A feasible solution must contain the correct total number of reserves (*R*) and have no more than *maxCA* scheduled to begin their duty at the departure time *d* (as no more than this will be required to cover crew absence at departure *d*).

The objective of the probabilistic crew absence model is to minimise cancellations due to crew absence plus the cancellation measure contributions due to reserve induced delay, as expressed by Equation 9. The objective when using the probabilistic crew absence model is to find a reserve crew schedule *X*, such that the cost of the resulting probabilities *P* and cancellation measures *CM* are minimised. For this research, the objective function can be expressed by Equation 9, which sums the cancellations plus reserve induced delay.

$$\text{ObjVal} = \sum_{d=1}^n \sum_{e=1}^{|p_d|} p_{d,e} + \sum_{d=1}^n \text{CM}_d \quad (9)$$

4.10 Improved Model

Early work with the *CAM* as presented so far identified a problem in that the *CAM* was underestimating the probabilities of cancellations due to crew absence. The reason for this is that the *CAM* implicitly assumes that the total number of absent crew is always exactly the expected value. However, in simulation the total number of crew that are absent in any given run of the simulation varies a great deal, and on bad days, once reserve crew have been used, cancellations spike. In fact, when all crew have equal probabilities of absence (as assumed in this work) the total number of absent crew follows a binomial distribution. For airline schedules involving many crew the Poisson distribution can be used as an approximation because it is the limiting case of the binomial distribution.

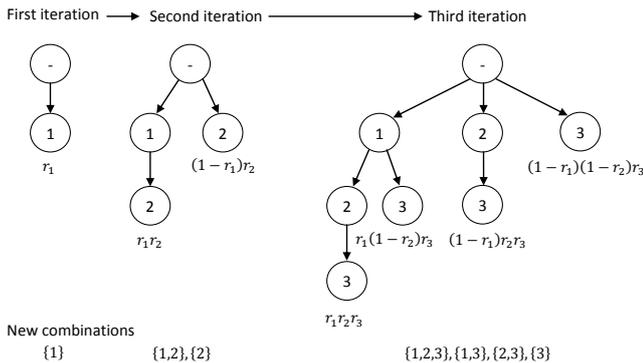


Figure 1: A growing reserve crew combination tree

Figure 2 illustrates this problem of underestimation by displaying the predicted cancellation probabilities derived from the *CAM* (purple) in comparison to those from repeat simulations (blue) for the example airline schedule described at the start of Section 5, with a reserve crew schedule, consisting of 12 individuals, derived from a greedy heuristic algorithm. The predictions from the *SPCAM* are also shown (cyan).

To remedy the cancellation underestimation problem, a refined probabilistic crew absence model (*CAM**) uses the distribution of the total number of absent crew (i.e. the binomial distribution) to evaluate reserve crew schedules simultaneously over a distribution of P matrices, denoted O , where matrix O_z corresponds to a Q matrix where the total expected number of absent crew is z . With z expected crew absences, n pairings, and up to $MaxCA$ crew absences per pairing, the probability of any given crew member being absent for a specific crew absence will be $p1 = \frac{z}{MaxCA \times n}$. The probability that there will be exactly z simultaneous crew absences is then given by $binomP(z; n \times MaxCA, p1)$. Equation 10 can then be used to obtain an improved estimation of the cancellation probabilities due to crew absence, where $eval(O_z, X)$ is a function to evaluate reserve crew schedule X using matrix O_z . The details of that function constitute Sections 4.4 through 4.8. This is identical to the approach used by the *CAM* (see Equation 9) except that the procedure is repeated over the distribution of Q matrices ($O_z \in O$) and a P matrix is calculated for each one.

Equations 9 and 10 are used as the objective functions in the search performed by the heuristic algorithms described in Section 5.2.

$$ObjVal^* = \sum_{z=1}^n binomP(z; n \times MaxCA, p1) \times eval(O_z, X) \quad (10)$$

For high values of z (total number absent) the associated binomial distribution probabilities become very small and can be ignored. For instance, the iteration could be stopped once the cumulative probability reaches a specific value, such as 0.95, 0.99 or 0.999. 0.999 was used as a stopping criterion in the experiments considered here. This approach vastly decreases the amount of time required for evaluating Equation 10, whilst maintaining the accuracy of the model. Validation experiments showed that accounting for more than the first 0.999 of the cumulative probability distribution of total number of absent crew resulted in no noticeable benefit in terms of prediction accuracy or the quality of the reserve crew schedule which was derived using the model.

Figure 2 also shows the results for the modified model (blue) and demonstrates that the *CAM** gives cancellation predictions of greater accuracy compared to the *CAM*, where the simulation predictions are treated as the target values.

5 Experimental Results

The models are used within a heuristic search to find a good reserve crew schedule. To test the validity of the model which has been developed, the cancellation probabilities for the resulting reserve crew schedules are compared with those from a Monte-Carlo simulation, which is the common method of evaluating schedules. Despite its popularity, the existing Monte-Carlo simulation approach has a potentially prohibitively long runtime. The aim of this work is two-fold: firstly, to find a model which will be a good proxy for the simulation results, but without the long

simulation time; and secondly, to find an effective method for off-line scheduling of reserve crew using this model. Two questions will therefore be answered here: how accurate is the cancellation prediction from the model in comparison to using a simulation approach; how effective is the reserve crew schedule which is found by the heuristic search in comparison to that which is found by other approaches.

The *SPCAM*, *CAM* and *SDM* are tested through experimentation. Two further models are also tested: *CAM** and *SDM**, which correspond to *CAM* and *SDM* respectively, with the addition of the model refinement of Section 4.10. The *SPCAM* implementation uses the probability of at least one crew absence affecting each departure as the single input probability for each departure. Apart from this the *SPCAM* is the same as the *CAM*.

The experiments are based on a real airline schedule. We note that the general results and conclusions in the following have been replicated in a variety of test instances including those from a random airline schedule generator (See Chapter 10 of Bayliss (2015)). We now focus the analysis by considering the (representative) results from a single realistic representative test instance. The test schedule is 2 days in length, with 283 departures from the hub station. There are 209 teams of crew and 74 aircraft covering a total of 566 flights. 140 of the crew teams begin their pairings at the hub station, these crew teams are subject to crew absence uncertainty. Each member of crew has a 1% chance of being absent. The 1% figure was derived from real airline crew absence data as a representative average crew absence rate. The proposed approach could also take as input all of the individual crew absence probabilities. Each team of crew consists of 4 members. There are 12 reserve crew available for scheduling. The aircraft routings are taken directly from real airline schedule data, the scheduled departure and arrival times are adjusted so that the scheduled block times are equal to the average actual block times. The journey time distributions for each origin-destination pair are derived from real flight data. The aircraft turn times and crew sit times are set to the minimum values. As a result of the preceding, the average delay risk in the test schedule was exactly 50% because the allocated block times were set so that each flight has a 50% chance of taking longer than the allocated flight time. This means that we have made the test instance more challenging in terms of the risk of delay by tightening them.

The input crew schedules were generated using a set partitioning model (described in Barnhart et al. (2003)) solved in CPLEX. The average rate of mid duty crew aircraft changes is 0.44. For full details of the test instance, see Bayliss et al. (2015).

The following experiments were implemented on a laptop with an 2.4GHz dual core Intel Core i7-5500U CPU, with 8Gb of RAM. All models, algorithms and the simulation were implemented in Java. The validation simulation reflects exactly the problem structure description of Section 3.

<i>SPCAM</i>	<i>CAM</i>	<i>CAM*</i>	<i>SDM</i>	<i>SDM*</i>	<i>SIM</i>
2.134	11.79	27.23	14.07	32.58	18192

Table 1: Comparison of evaluation times (in milliseconds) of different versions of the probabilistic model and 20000 repeat simulations

Table 1 displays the time in milliseconds for a single evaluation of each variant of the probabilistic model (the first five columns) compared with the time required to run 20000 repeat simulations (labelled *SIM*). In order to determine an appropriate number of repeat simulations for comparison, the convergence of a 10-fold cross validation (Hastie et al., 2008) was analysed. The average cancellation measure RMSE, converges to approxi-

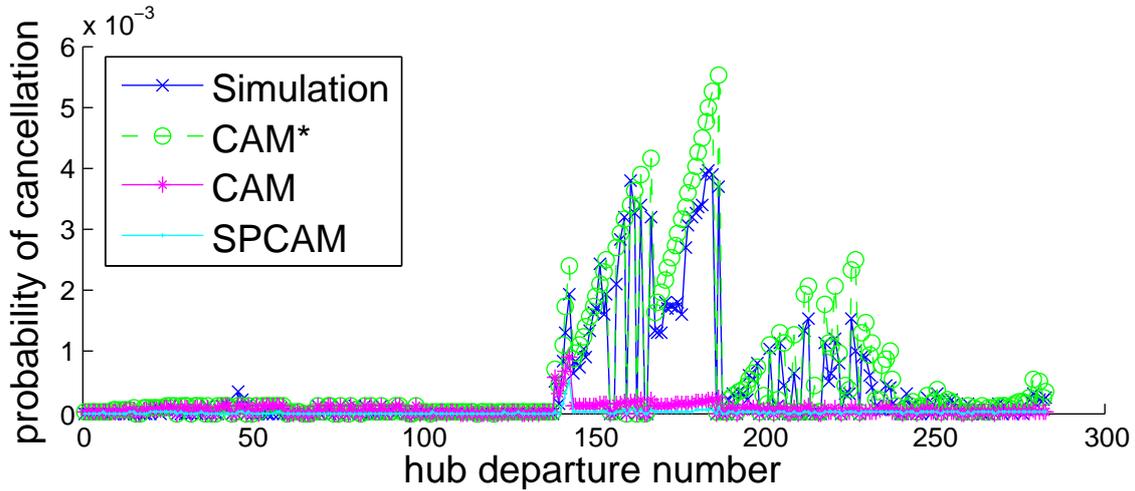


Figure 2: Cancellation predictions from evaluations of the *SPCAM*, *CAM* and refined probabilistic crew absence model (*CAM**) compared to those derived from repeat simulations

mately 0.004, for a simulation sample size of 20000. The conclusion here is that the proposed model offers significant speed benefits in comparison to the simulation based evaluation method. Additionally the probabilistic model approach circumvents the problem of noise associated with simulation based approaches.

5.1 The Effect Of The Value Of $delExp$

The *SDM* and *SDM** both include a penalty term for reserve induced delays, as shown in Equation 5. In order to understand the influence of the value of $delExp$ in the penalty term, reserve crew schedules were derived using a simulated annealing algorithm (see Section 5.2) for each value of $delExp$ from 0 to 10 in increments of 1, with 10 repeats for each. Each was then tested in 20000 repeat simulations to derive the associated cancellation rate and the average delay. Figure 3 shows the effect that varying the value of $delExp$ has on the cancellation rate and the average delay.

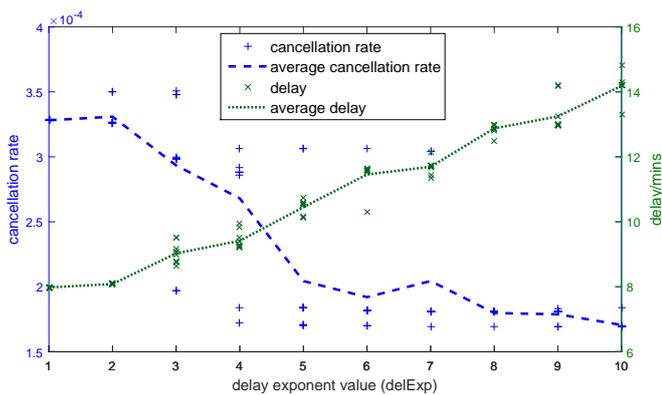


Figure 3: The effect of the value of $delExp$ used when scheduling reserve crew on the (per flight) cancellation rate and the average delay

Figure 3 shows how the average delay increases and the cancellation rate decreases as $delExp$ is increased. The choice of the value of $delExp$ requires a decision maker input, to select an appropriate trade off between cancellation minimisation and delay minimisation. In the following $delExp = 2$ is assumed. In general $delExp > 1$ is advised because multiple delays that sum to *CT* are assumed to be preferable to an actual cancellation, and

because a large delay is considered more damaging than a number of small delays that sum to the same large delay.

5.2 Experimental Design And Heuristic Search

As previously mentioned, the primary aim of this research is to find an effective surrogate for the lengthy simulation process, and the secondary aim of this research is to evaluate the feasibility of utilising the proposed model within a search for effective schedules. For this latter purpose, the *SPCAM*, *CAM*, *CAM**, *SDM* and *SDM** are now all used within a variety of different heuristic searches, attempting to derive good reserve crew schedules. The following heuristics are considered:

GH: The greedy heuristic adds reserve crew one at a time to a reserve crew schedule, each time selecting the start time that reduces the objective function the most, continuing in this fashion until all of the reserve crew have been scheduled.

LS: Local search starts from a randomly generated initial solution (randomly generated start time indices). In each iteration all solutions neighbouring the incumbent solution are evaluated, the solution which reduces the objective value the most is accepted. If no improving solution is available the algorithm terminates. Local search uses the cut-and-insert neighbourhood structure, considering all potential solutions that have one reserve start time different to that of the incumbent solution.

GH+LS: LS starting from the GH solution.

SA: The simulated annealing Kirkpatrick et al. (1983) implementation uses the cut-and-insert neighbourhood. Each iteration randomly selects a neighbouring solution, which is accepted if it is an improving move. A non-improving move is accepted with probability $e^{-\Delta/T}$. Δ is the increase in the objective value associated with the non-improving move, T is the current temperature. The cooling scheme (value of T at any given iteration) is based on an exponential decay starting from T_0 equal to the maximum number of hub departures in a crew pairing and reaching a final temperature of 0.000001 after 20000 iterations. The cooling scheme is a function of the number of evaluations that have been performed so far. Given that the exponential decay cooling scheme is fully defined by T_0 and the final temperature, these are the only free parameters. In this case they have been set deterministically so that only in the first iteration is the largest possible increase in cancellation measure in the cut and insert neighbourhood accepted with probability 1. We refer to this as

the *boiling point* initial temperature.

GA1: The first genetic algorithm Goldberg (1989) implementation uses a population size of 50, uses nine competitor tournament selection, a mutation rate of 0.01, single point crossover which is applied with probability 1 and all parent chromosomes are replaced with children chromosomes in each generation. These parameters were the best on average in a full factorial experiment of various populations sizes, mutation rates and tournament sizes. The average cancellation measures associated with each parameter value are given in Tables 2, 3 and 4. The genetic algorithm returns the best solution found after 20000 function evaluations. For best results the genetic algorithm parameters have to be reconsidered with each new application of the approach.

GA2: The second genetic algorithm implementation is the same as GA1, except that the mutation operator is replaced with a single iteration of the SA algorithm (only one random neighbouring solution is considered), which is applied to each member of the population, in each generation. The SA uses the same temperature scheme as the SA algorithm. GA2 is limited to a total of 20000 evaluations, which means that GA2 uses half as many generations as GA1, the other half are used evaluating SA generated mutations. GA2 is similar to a memetic algorithm (Hart et al., 2005), because of the addition of a local search based approach, to the algorithm. The main difference being that, because only one iteration of simulated annealing is used, the approach is actually closer to a mutation operator than an application of local search.

The parameters for the cooling scheme of the SA mutation steps are the same as those for SA (boiling point initial temperature), but in this case the temperature drops twice as fast because the computational budget is shared equally between the GA and SA aspects of this method.

All heuristics except for the GH are limited to 20000 function evaluations. The GH only requires around 3000 function evaluations to derive a reserve crew schedule for the given problem. 20000 function evaluations take around 10 minutes on the above described hardware/software combination. This approach enables a fair test of the different heuristics in a situation where solutions are required within a set time limit. Furthermore a fixed computational budget increases the comparability of the results derived from each solution methodology. Each heuristic is then repeated 10 times using each probabilistic reserve crew schedule evaluator (*SPCAM*, *CAM*, *CAM**, *SDM* and *SDM**).

For validation purposes each derived reserve crew schedule is then tested in 20000 new repeat simulations to derive cancellation and delay based performance measures. Each repeat simulation uses different stochastic inputs to instantiate numbers of absent crew for each crew pairing. There are two main reasons for the simulation testing of the reserve crew schedules derived above. Firstly, it will demonstrate that the improved probabilistic models (*CAM* and *CAM**) provide accurate cancellation probabilities for individual flights (Section 5.3), and secondly this will confirm that the use of the probabilistic models as surrogate objective functions has been effective (Section 5.4).

5.3 Cancellation Prediction Accuracy

This section compares the predicted average cancellations due to crew absence from the *SPCAM*, *CAM* and *CAM** with the average cancellation rates observed in repeat simulations. The aim is to verify whether the predictions from the models match those from the simulation approaches. These experiments were repeated for each reserve crew schedule derived from the exper-

iment described above in Section 5.2. Note that the cancellation predictions of the *SDM* and *SDM** match those of the *CAM* and *CAM** respectively.

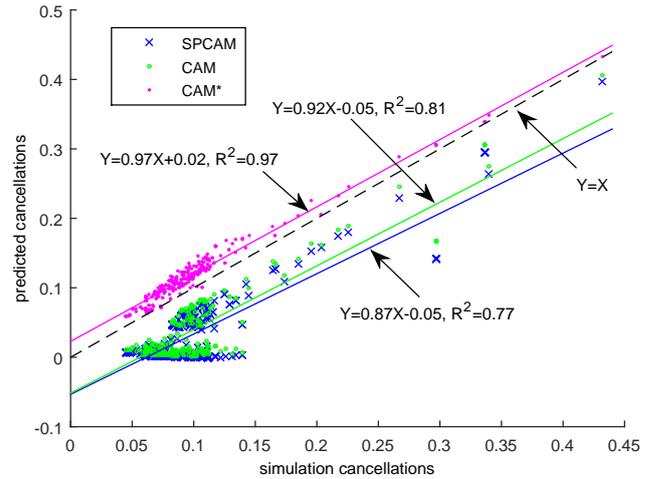


Figure 4: Cancellation predictions from evaluations of the *SPCAM*, *CAM* and *CAM** compared to those derived from repeat simulations

Figure 4 confirms that the *CAM* (and the *SPCAM*) underestimates cancellations due to crew absence, and that the *CAM** successfully alleviates this problem. The *CAM** does however systematically overestimate cancellations due to crew absence, each time in a manner similar to that demonstrated in Figure 2. Possible reasons for this include: 20000 repeat simulations are not enough to capture a representative sample of the worst case scenarios in which cancellation spikes occur; cumulative rounding errors; the probabilities of reserve combinations calculated based on filters 1 and 2 of Section 4.8 have additional factors/intricacies which have not yet been uncovered. Despite this, the *CAM** consistently gives the most accurate cancellation predictions, which is supported by the linear trend equation, which has a gradient close to one, an intercept close to zero and a high correlation coefficient compared to those of the *SPCAM* and the *CAM*. On the plus side, overestimations of cancellations due to crew absence are not as potentially damaging as underestimations, because cancellations are the most severe outcome from crew absences. A systematic overestimation of cancellations due to crew absence could actually be beneficial, as this corresponds to a more risk averse approach to reserve crew scheduling.

5.4 Reserve Crew Scheduling Application

The results of the experiment described in Section 5.2 are now used to show the effects that the probabilistic models and the search heuristics which were used to schedule reserve crew have on the quality of the resultant reserve crew schedules. The average cancellation measures (cancellations plus cancellation measures of delay) derived for each reserve crew schedule tested in 20000 repeat simulations are used as the measures of reserve crew schedule quality. Figure 5 shows the average cancellation measure of the best reserve crew schedules from 10 repeats of each heuristic used in conjunction with each evaluator. The results show that as the evaluator complexity increases (*SPCAM* to *SDM**) the average cancellation measure decreases - i.e. the approach is able to better utilise the reserves to avoid cancellations.

Table 5 shows the effectiveness of the different approaches to crew scheduling, illustrating for each approach the cancella-

Population size	10	20	50	100	200	500	1000
Average cancellation measure	0.3127	0.3046	0.2997	0.2997	0.3032	0.3168	0.3510

Table 2: Population size effect on average cancellation measure

Mutation rate	1	0.5	0.1	0.05	0.01	0.005	0.001	0.0005	0.0001
Average cancellation measure	0.3119	0.3150	0.3096	0.3133	0.3094	0.3125	0.3145	0.3148	0.3115

Table 3: Mutation rate effect on average cancellation measure

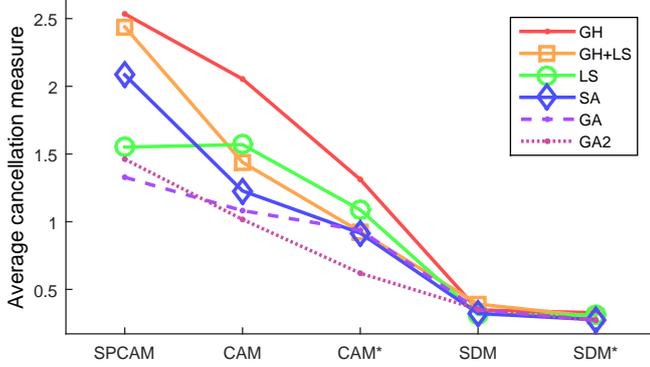


Figure 5: The effect of each variant of the probabilistic crew absence model and the search heuristics used to schedule reserve crew on the resultant simulation derived average cancellation measure

tion measure results, the average number of cancellations due to (uncovered) crew absence and excessive delay, the average size of the delays which are greater than the delay threshold, the probability of having a delay over the threshold, the rate of reserve crew utilisation and the maximum cancellation measure over 20000 repeat simulations. The *CAM** and *SDM** evaluators typically lead to respectively higher quality reserve crew schedules than the *CAM* and *SDM* evaluators. This result is supported by the results given in Table 5, which validates the model refinement of Section 4.10.

The variation in Figure 5 between the average cancellation measures for reserve crew schedules derived from the *SPCAM*, *CAM* and *CAM** evaluators can be explained by them not allowing for reserve induced delay. This is confirmed by the results of Table 5.

The low variation in Figure 5 between the results for the *SDM* and *SDM** evaluators can be explained by the fact that each heuristic is capable of deriving a good quality solution, provided that the evaluator used includes all of the aspects on which solution quality is judged. For the *SPCAM*, *CAM* and *CAM** evaluators, GA2 always gave the best reserve crew schedule. For the *SDM* evaluator, GA1 gave the best solution. For the *SDM**, SA gave the best (overall) reserve crew schedule.

The LS approach did not attain any of the best reserve crew schedules found. An explanation for this is that the cut-and-insert neighbourhood structure had a size of 3396 neighbouring solutions at any given iteration, which all had to be evaluated before accepting the best neighbouring solution. This meant that the LS approach never reached a local optimum, because only 5 full iterations could be performed within the 20000 evaluations limit. For this reason a Tabu Search was not implemented, as it would not have had the chance to exploit a tabu list.

Figure 5 also implies that the accuracy of the evaluator is, in general, more important than the complexity of the search algorithm used to derive a reserve crew schedule.

5.5 Comparison With Alternative Approaches

In this section the best reserve crew schedules derived using each variant of the probabilistic crew absence model are compared with each other and with a range of alternative approaches to reserve crew scheduling. The alternative approaches considered are as follows:

USR: The uniform start rate heuristic schedules the available reserve crew at times corresponding to equal intervals of hub departures. E.g. if there are 25 hub departures and 5 reserve crew, $X = \{1, 6, 11, 16, 21\}$.

AREA: The area under the graph approach uses repeat simulations to determine the cumulative demand for reserve crew at each hub departure. Then, the available reserve crew are scheduled at equal intervals of cumulative demand, or equivalently at intervals of equal areas under the demand graph. The *AREA* approach was also used in Bayliss et al. (2013) and Bayliss et al. (2017).

MIPSSM: The mixed integer programming simulation scenario model Bayliss et al. (2017) was described in Section 2. In contrast to the probabilistic approaches developed in this work, this method represents a scenario-based approach to solving the same problem. For each repeat, the *MIPSSM* is solved using 20 randomly generated input scenarios.

SSH: The scenario selection heuristic Bayliss et al. (2017) is a heuristic approach for selecting the scenarios which are included in the *MIPSSM* formulation. The scenario selection heuristic is based on adding the scenario which increases the objective value of the *MIPSSM* the most, then resolving the *MIPSSM* to obtain a more robust reserve crew schedule. This process continues until no new scenario can be found which increases the objective value of the *MIPSSM* more than the single largest objective contribution of any scenario which is already included in the model.

The results in Table 5 correspond to the best of 10 repeats of each of the above described alternative approaches. These are compared with the best reserve crew schedules derived using each of the probabilistic evaluators considered in this work. Note that when the best single repeats are replaced with the average of the 10 repeats of each method, the ordering of the methods, in terms of average cancellation measure, is the same. The average cancellation measure for the best repeats of each method are given to indicate the potential of each approach. The standard deviation from the 10 repeats indicates the reliability of each method.

In Table 5 the average cancellation measures show that the *SDM** results in the lowest average cancellation measure. The *SSH* also attains a very low average cancellation measure, however the variance, as measured by the standard deviation of the repeats for the scenario-based approaches, was much greater than that of the *SDM**, which is because the scenario-based approaches only ever have an incomplete picture of crew absence uncertainty, which is that captured by a limited number of input scenarios (See Bayliss et al. (2017) for more details). In comparison, the *SDM** is much more reliable across repeat experiments.

Tournament size	2	3	4	5	6	7	8	9	10
Average cancellation measure	0.3570	0.3227	0.3129	0.3120	0.3040	0.3040	0.3006	0.2980	0.3015

Table 4: Tournament size effect on average cancellation measure

Method measure type	Method name	Average Cancellation measure		Average cancellations		Delays > delay threshold		Reserve utilisation rate	Maximum cancellation measure
		best repeat	St dev (10 repeats)	due to absence	due to delay	Average (minutes)	probability		
-	no res	10.1668	0	10.0991	0	6.802	0.0818	0	37.02
Probabilistic	<i>SPCAM</i>	1.3282	0.4078	0.0882	0.0122	17.293	0.1073	0.4630	20.06
	<i>CAM</i>	1.0167	0.2573	0.0712	0.0112	15.326	0.1050	0.4629	19.41
	<i>CAM*</i>	0.6196	0.3754	0.0515	0.0066	12.186	0.1005	0.4627	19.07
	<i>SDM</i>	0.3147	0.1057	0.0958	0.0010	8.574	0.0942	0.4588	12.35
	<i>SDM*</i>	0.2720	0.0124	0.1026	0.0006	8.019	0.0933	0.4577	15.45
Heuristic	<i>USR</i>	1.3963	0	0.2967	0.00785	16.535	0.102	0.4543	19.17
Simulation	<i>AREA</i>	0.9823	0	0.3366	0.0061	12.44	0.0983	0.4378	14.48
Scenario based	<i>MIPSSM</i>	0.4452	0.0899	0.1452	0.0006	9.191	0.0928	0.4555	13.37
	<i>SSH</i>	0.2953	0.1161	0.096	0.0032	8.241	0.0932	0.458	15.55

Table 5: Comparison of all approaches to reserve crew scheduling considering additional performance measures

Table 5 gives the average expected number of cancellations due to crew absence and due to delays exceeding the cancellation threshold, for each method. When no reserve crew are scheduled, there are an average of 10 cancellations, all of which are due to absence. This means that all of the observed cancellations due to delay, in Table 5, are caused by reserve induced delays which have been propagated and have caused delays above the cancellation threshold later on.

Cancellations due to delay are highest for the *SPCAM*, which is because this approach does not penalise reserve induced delays. The *SDM* and *SDM** reduce cancellations due to delay, because they do penalise reserve induced delay. The *MIPSSM* minimised cancellations due to delay, but not the average cancellation measure, which is its objective. The *CAM** achieved the lowest cancellations due to absence, which can be attributed to it overestimating cancellations due to absence, resulting in reserve crew schedules which are highly risk averse in terms of this type of disruption.

The delay based performance measures show that the lowest results for delays above the delay threshold and their probabilities of occurring were achieved by the approaches which penalise reserve induced delay, i.e. *SDM* and *SDM**.

The reserve utilisation rate results show that reserve utilisation rate is not an indicator of the quality of a reserve crew schedule, because the maximum reserve utilisation rate (that of the *CAM*) also attained a relatively high average cancellation measure. This means that it is possible to obtain worse results even when more reserves are allocated, if these reserves are allocated poorly.

It is interesting to note that the *SDM** attains the lowest average cancellation measure by finding a balanced trade off between cancellations due to absence and reserve-induced delays, i.e. it trades cancellations due to absence for reduced delays and cancellations due to delays.

The maximum cancellation measure performance attribute (last column) gives the worst case total cancellation measure from the 20000 repeat simulations. When no reserve crew were scheduled, a total cancellation measure of 37 was accumulated in the worst case. The *MIPSSM* has the lowest maximum cancellation measure, which may be a particular strength of the scenario-based approaches.

In terms of a comparison between the different types of approaches to reserve crew scheduling, the probabilistic and

	Reserve use	Cancellation	Crew swap	Aircraft swap
<i>SDM</i>	0.019435	0.000326	0.001788	0.001825
<i>SSH</i>	0.019422	0.000350	0.001690	0.001824
<i>AREA</i>	0.018565	0.001211	0.002200	0.001869
<i>USR</i>	0.019266	0.001076	0.002578	0.001850

Table 6: Comparison of recovery action rates (per hub departure) for the 4 main types of reserve crew schedule (20000 repeat simulations)

scenario-based approaches are the strongest. The results given here suggest that the *SDM** has the edge on the *SSH* approach, however future improvements of the scenario-based approaches could possibly change this. In fact, each of these general approaches has their own unique strengths and weaknesses. Probabilistic approaches model all possible outcomes as a single scenario, in which all disruption events have some probability of occurring. This approach may not be able to capture the intricacies associated with any particular scenario, which is the strength of the scenario-based approaches. The weakness of the scenario-based approaches is that the models become intractable very quickly as the number of input scenarios is increased. As a result, scenario-based approaches only ever have access to a limited sample of all possible scenarios, which is the strength of the probabilistic approaches.

Table 6 shows the recovery action rates per hub departure for the best approach from each of the four main types of reserve crew scheduling approach (as indicated by the results from Table 5). It shows that the more effective approaches reduce the rate at which crew swap recovery actions are required. The explanation for this is that crew schedules are more time-constrained than the aircraft routings. In particular a poor reserve crew schedule will cause delays which propagate and increase the need for crew swaps later on.

Figure 6 displays the (\log_{10}) average cancellation measures from 20000 repeat simulations of the best reserve crew schedules derived from each of the main approaches to reserve crew scheduling. It displays the similarity in the quality of reserve crew schedules derived from the *SDM* and *SSH* approaches (although *SDM* is the more reliable method, see column 4 of Table 5). Each percentile data series features three points of inflexion,

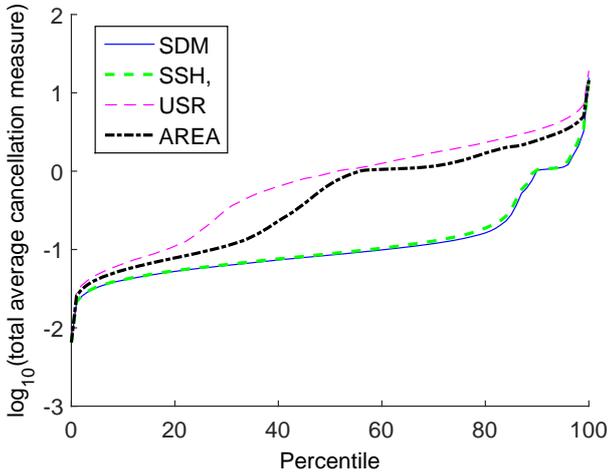


Figure 6: Average (\log_{10}) cancellation measure percentiles for the best reserve crew schedules from each of the main approaches to reserve crew scheduling

these corresponding to: best case performance (low average cancellation measures); worst case performance measures; and the spike in cancellation measures referred to in Section 4.10, which is observed more often in the poorer reserve crew schedule approaches.

6 Conclusion

This work has described and evaluated a new mathematical model for reserve crew scheduling, which is an expansion and modification of a model which was previously introduced by the authors. The improved model improves the accuracy of the earlier model by accounting for: the possibility of multiple crew being absent simultaneously from crew teams; the structure of crew pairings in terms of the effect that these have on the dependencies of the probabilities of cancellations between flights associated with the same crew team; and the potential for reserve induced delay. The improved model required an algorithm for enumerating feasible combinations of reserve crew for all possible crew absence disruptions and a function for converting reserve induced delays into a measure of cancellation. The resultant model, was found to underestimate cancellations due to crew absence, which was found to be the result of not accounting for the variance in the total number of crew that can be absent on the day of operation. A model refinement allowed for this effect, which led to more accurate predictions.

Each variant of the probabilistic crew absence model was used as the evaluator in a range of heuristic search algorithms, seeking good reserve crew schedules. The results were used to confirm the increased prediction accuracy of the model refinement and also to find the best reserve crew schedules which could be derived from each evaluator. The reserve crew schedules derived from the *SDM** gave the best overall results, thus validating each of the model improvements proposed in this work. It was found that the success of the *SDM** approach owed to it finding a trade off between cancellations due to absence and reserve induced delays. The *SDM** accepted a slightly increased risk of cancellation due to crew absence in return for significant reductions in average delays and cancellations due to delays. The approaches were also compared to a variety of alternative approaches, including scenario-based approaches, one of which gave results of comparable, although slightly lower quality. It was concluded that

probabilistic and scenario-based approaches, are polar extremes in terms of their respective strengths and weaknesses.

This work focussed on a single fleet, crew rank and crew qualification example. Future work could involve modifying the current model to apply to the case where there are multiple fleet types, crew ranks and qualifications. To do this, Algorithm 3 would have to generate combinations of feasible reserve crew involving crew of a variety of ranks and qualifications. This is because crew qualifications can overlap with respect to the fleet types they are qualified to operate on. Additionally, high rank crew have the possibility of flying-below-rank, so any given combination of reserve crew has the potential to be used to cover a range of different-rank combinations of absent crew. For a two crew rank example, the probability matrix element $p_{d,e}$ would become $p_{d,e,f}$, to correspond to the probability that e low rank crew and f high rank crew are unavailable for flight d . An additional consequence of considering a multiple fleet, crew rank and qualification example, would be that the assumed earliest reserve start time order policy would have to be able to differentiate between reserve crew with equal start times but different rank and qualification combinations. One possibility would be to order reserve crew according to earliest start time, and break ties according to the preferred rank and qualification for the given disruption. If flying-below-rank is deemed undesirable, high rank reserve crew could be considered last for covering low rank absent crew.

Additional future work includes the potential use of the proposed model as an on-line decision tool. In order to implement the approach in an on-line context the model would require, as input, a snapshot of the current schedule in addition to the most recent information regarding the crew who are known to be absent. One of the main complexities of an on-line implementation is the integration of the model into an airline's existing decision support systems. This is complex because each new method has its own set of required inputs and outputs. Mathaisel (1996) provides an account of how such complexities can be addressed by integrating recovery for the different layers (passenger, crew and aircraft) of an airline's schedule. In order to evaluate a single reserve crew use decision (compared to holding the reserve crew for later use) two evaluations of the model would be required. In the current implementation evaluations can be performed within a fraction of a second. However evaluation time increase linearly with the length of the scheduling horizon that is considered and exponentially as the number of reserve crew increase. There are many useful approximations that can reduce this problem, including truncating calculations for reserve crew who have a below-threshold probability of remaining available or being used in the first place. One such approximation was considered in Section 4.10.

In this work we have focussed on minimising absence disruption propagation through reserve crew scheduling. Reserve induced delays were taken into account but delay propagation was not modelled. In general, delay propagation increases throughout the day as disruptions from many different sources accumulate, including from weather, ground and air congestion, unscheduled maintenance requirements and crew absence. Future work will combine an explicit delay propagation model with the model presented here, and evaluate the benefits for the resulting reserve crew schedules of doing so.

7 Acknowledgements

The authors would like to thank the EPSRC LANCS initiative (grant ref EP/F033214/1) for funding this research.

REFERENCES

- Abdelghany, A., Ekollu, G., Narasimhan, R., and Abdelghany, K. (2004). A proactive crew recovery decision support tool for commercial airlines during irregular operations. *Annals of Operations Research*, 127:309–331.
- Akartunal, K., Boland, N., Evans, I., Wallace, M., and Waterer, H. (2013). Airline planning benchmark problems part i: Characterising networks and demand using limited data. *Computers Ops Res*, 40:775–792.
- Amita, J., Singh, J. S., and Kumar, G. P. (2015). Prediction of bus travel time using artificial neural network. *International journal of traffic and transport engineering*, 5(4):410–424.
- Barnhart, C., Cohn, A. M., Johnson, E. L., Klabjan, D., Nemhauser, G. L., and Vance, P. H. (2003). Airline crew scheduling. *Handbook of Transportation Science*, pages 517–560.
- Bayliss, C. (2015). *Airline reserve crew scheduling under uncertainty*. PhD thesis, University of Nottingham.
- Bayliss, C., De Maere, G., Atkin, J. A. D., and Paelinck, M. (2012). Probabilistic Airline Reserve Crew Scheduling Model. In Dellling, D. and Liberti, L., editors, *12th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*, volume 25 of *OpenAccess Series in Informatics (OASISs)*, pages 132–143, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- Bayliss, C., De Maere, G., Atkin, J. A. D., and Paelinck, M. (2015). Test instance data for section 5. www.asap.cs.nott.ac.uk/external/atrbenchmarks/index.shtml. last accessed October 2017.
- Bayliss, C., De Maere, G., Atkin, J. A. D., and Paelinck, M. (2017). A simulation scenario based mixed integer programming approach to airline reserve crew scheduling under uncertainty. *Annals of Operational Research*, 252(2):335–363.
- Bayliss, C., De Maere, G., Atkin, J. A. D., and Paelinck, M. (2013). Scheduling airline reserve crew to minimise crew related delay using simulated airline recovery and a probabilistic optimisation model. In *proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 1944–1950.
- Burke, E. K., Causmaecker, P. D., De Maere, G., Mulder, J., Paelinck, M., and Berghe, G. V. (2010). A multi-objective approach for robust airline scheduling. *Computers and Operations Research*, 37:822–832.
- Chang, S.-C. (2012). A duty based approach in solving the air-crew recovery problem. *Journal of Air Transport Management*, 19:16–20.
- Deng, W., Chen, R., He, B., Liu, Y., Yin, L., and Guo, J. (2012). A novel two-stage hybrid swarm intelligence optimization algorithm and application. *Soft Computing*, 16(10):1707–1722.
- Deng, W., Zhao, H., Liu, J., Yan, X., Li, Y., Yin, L., and Ding, C. (2015). An improved caco algorithm based on adaptive method and multi-variant strategies. *Soft Computing*, 19(3):701–713.
- Deng, W., Zhao, H., Yang, X., Xiong, J., Sun, M., and Li, B. (2017a). Study on an improved adaptive pso algorithm for solving multi-objective gate assignment. *Applied Soft Computing*, 59:288–302.
- Deng, W., Zhao, H., Zou, L., Li, G., Yang, X., and Wu, D. (2017b). A novel collaborative optimization algorithm in solving complex optimization problems. *Soft Computing*, 21(15):4387–4398.
- Dillon, J. E. and Kontogiorgis, S. (1999). US airways optimises the scheduling of reserve flight crews. *Interfaces*, 29(5):123–131.
- Duck, V., Ionescu, L., Kliewer, N., and Suhl, L. (2012). Increasing stability of airline crew and aircraft schedules. *Transportation Research Part C*, 20:47–61.
- Gaballa, A. (1979). Planning callout reserves for aircraft delays. *Interfaces*, 9(2):78–86.
- Goldberg, D. E. (1989). *Genetic algorithms in search optimization and machine learning*. Addison-Wesley, Reading, Massachusetts, Menlo Park, California, Sydney, Don Mills, Ontario, Madrid, San Juan, New York, Singapore, Amsterdam, Wokingham, England, Tokyo, Bonn.
- Gu, B. and Sheng, V. S. (2017). A robust regularization path algorithm for v-support vector classification. *IEEE Transactions on Neural Networks and Learning Systems*, 28(5):1241–1248.
- Gu, B., Sheng, V. S., Tay, K. Y., Romano, W., and Li, S. (2015). Increment support vector learning for ordinal regression. *IEEE Transactions on Neural, Networks and Learning Systems*, 26(7):1403–1416.
- Hart, W. E., Krasnogor, N., and Smith, J. E. (2005). *Recent advances in memetic algorithms*. Springer-Verlag, Berlin Heidelberg.
- Hastie, T., Tibshirani, R., and Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, chapter 7. Springer Series in Statistics. Springer, second edition.
- Kirkpatrick, S., Gelatt, C. D., Jr, and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.
- Mathaisel, D. F. X. (1996). Decision support for airline system operations control and irregular operations. *Computers Ops Res*, 23(11):1083–1098.
- Paelinck, M. (2001). KLM cabin crew reserve duty optimisation. In *AGIFORS proceedings*.
- Peterson, J. D., Solveling, G., Johnson, E., Clarke, J.-P., and Shebalov, S. (2012). An optimisation approach to airline integrated recovery. *Transportation Science*, 46(4):482–500.
- Rosenberger, J. M., Johnson, E. L., and Nemhauser, G. L. (2003). Rerouting aircraft for airline recovery. *Transportation Science*, 37(4):408–421.
- Schaefer, A. J., Johnson, E. L., Kleywegt, A. J., and Nemhauser, G. L. (2005). Airline crew scheduling under uncertainty. *Transportation Science*, 39(3):340–348.
- Sohoni, M., Lee, Y.-C., and Klabjan, D. (2011). Robust airline scheduling under block-time uncertainty. *Transportation Science*, 45(4):451–464.

- Sohoni, M. G., Johnson, E. L., and Bailey, T. G. (2006). Operational airlines reserve crew scheduling. *Journal of Scheduling*, 9:203–221.
- Weide, O., Ryan, D., and Ehrgott, M. (2010). Iterative approach to robust aircraft routing and crew scheduling. *Computers and Operations Research*, 37:833–844.
- Wu, C.-L. (2010). *Airline operations and delay management*. Ashgate, Burlington, USA and Farnham, England.
- Xue, Y., Jiang, J., Zhao, B., and Ma, T. (2018). A self-adaptive artificial bee colony algorithm based on global best for global optimization. *Soft Computing*, 22:2935–2952.
- Yen, J. W. and Birge, J. R. (2006). A stochastic programming approach to the airline crew scheduling problem. *Transportation Science*, 40(1):3–14.