AIRLINE DISRUPTION RECOVERY USING SYMBIOTIC SIMULATION AND MULTI-FIDELITY MODELLING

Luke Rhodes-Leader
STOR-i CDT
Lancaster University, United Kingdom
l.rhodes-leader@lancaster.ac.uk

Dr. Bhakti Stephan Onggo
Trinity Business School
Trinity College Dublin
Dublin, Ireland
onggos@tcd.ie

Dr. David J. Worthington
Department of Management Science
Lancaster University Management School
Lancaster University, United Kingdom
d.worthington@lancaster.ac.uk

Prof. Barry L. Nelson
Department of Industrial Engineering & Management Sciences
Northwestern University
Evanston, IL 60208 USA
nelsonb@northwestern.edu

ABSTRACT
The airlines industry is prone to disruption due to various causes. Whilst an airline may not be able to control the causes of disruption, it can reduce the impact of a disruptive event, such as a mechanical failure, with its response by revising the schedule. Potential actions include swapping aircraft, delaying flights and cancellations. This paper will present our research into how symbiotic simulation could potentially be used to improve the response to a disruptive event by evaluating potential revised schedules. Due to the large solution space, exhaustive searches are infeasible. Our research is investigating the use of multi-fidelity models to help guide the search of the optimisation algorithm, leading to good solutions being generated within the time constraints of disruption management.

1 INTRODUCTION
One of the major problems faced by airlines is disruption to their schedules. Although a great deal of time and money is spent on preparing an optimal schedule, it is very rare that a flight programme will be carried out as intended within the operation. This can be caused by a variety of issues such as weather conditions and aircraft failures. This paper will focus on disruption caused by aircraft failures requiring unplanned maintenance, leading to problems with fleet allocation. The impacts of such an event can propagate through the system causing many more delays and cancellations, particularly if the airline has
a high aircraft utilisation. The response of an airline seeking to manage the burden of a disruptive event can have a large effect on the outcome. Each action corresponds to revising the schedule in some way, whether it be delaying or cancelling flights or exchanging aircraft.

The airlines industry is a very complex environment and it can be difficult to see what consequences a change in schedule may have. To help with the Aircraft Recovery Problem (ARP), a variety of approaches using mathematical programming techniques (such as integer programming) have been proposed. These can often deal with the complexity involved in a deterministic model. However, the airlines industry is also highly stochastic. The turn-around times, times spent queueing at airports and other possible failures are stochastic, meaning that further disruptions could occur during the recovery period. The ability of standard deterministic programmes to incorporate this stochasticity is limited, particularly as the number of sources of variation is large. The recovery problem in which variability is taken into account is known as the Stochastic Aircraft Recovery Problem (SARP). Only probabilistic models can achieve the levels of detail required. Unfortunately, due to the complexity, analytical models are rarely tractable.

Therefore, simulation seems to be a natural way of modelling airline operations. Furthermore, it is important that the model remains a good representation of the physical process by adapting its input parameters as the system evolves with time.

Symbiotic simulation was first proposed in Fujimoto et al. (2002) as a means of allowing a simulation system to interact with the physical system it models in a mutually beneficial way. The exchange of information between the two systems has the potential to improve both the performance of the physical system and the representation of the simulation model by adapting it to new circumstances. At the time of triggering, multiple strategies can be tested using the simulation model and the output used to find a good solution. Part of its attraction is its ability to be reused in new situations (either through different initial conditions or by a change in the underlying process) without requiring major programming alterations. Symbiotic simulation has received much interest in the areas of, for example, manufacturing and traffic management.

We believe that the abilities of symbiotic simulation have great potential within airline operations, particularly as a decision support system using reactive triggering to help tackle the SARP.

However, high-fidelity simulation models have non-negligible computation time, which proves problematic for search and optimisation algorithms when there is a large solution space and tight time constraints such as in airline disruption management. The use of the simulation must be selective in order to find good solutions within a reasonable time. Multi-fidelity modelling is the use of a low-fidelity model to quickly identify potentially good solutions that can then be simulated to give a more thorough evaluation of the solution’s performance. The research outlined in this paper is investigating the use of multi-fidelity modelling within the aircraft recovery problem.

This paper is organised as follows. A brief review of some of the approaches to disruption management in airlines operations and symbiotic simulation applications is in Section 2. In Section 3, we provide an overview of our current simulation and integer programming models for the SARP as well as how these mixed fidelity models are combined. A demonstration and preliminary results are given in Section 4. Following this, we discuss some of the future directions of the research.

2 LITERATURE REVIEW

There are many applications of deterministic optimisation within the literature to problems involving adapting an airline’s schedule. These include work by Rosenberger et al. (2003) who propose an integer programme based on a set packing formulation to assign aircraft to rotations whilst minimising the cost and a heuristic to help reduce the problem size. Love et al. (2005) formulate the problem using a network in which both aircraft and flights are represented by nodes and arcs represent the assignments and routes. Again, cost is the main objective. The authors describe a Steepest Ascent Local Search heuristic to find good solutions. However, the consequences of a disruption are random events and so are difficult to truly incorporate into a deterministic regime. Zhu et al. (2015) include some stochastic elements by using a stochastic programme
to deal with different potential times at which an aircraft truly becomes available. However, other stochastic elements are not considered. This is the major limitation of traditional optimization techniques.

Disruption handling has been considered within simulation before, particularly in the SimAir software initially proposed by Rosenberger et al. (2000). These included a simulation of an airline’s operation, evaluating different recovery strategies using deterministic tests and implementing the best decision under a particular disruptive event (Rosenberger et al. 2002). Lee et al. (2003) use SimAir to explore the robustness of a given schedule and a variety of response policies to disruption. At each event, the schedule is checked to see if any legal regulations, such as crew flying hours, are to be violated in subsequent flights. If so, a recovery module selects a recovery action from a list of heuristic responses, such as delaying flights and short cycle cancellations. The main aim of the recovery module is to resolve any illegalities detected, rather than to minimise the impact. However, the main focus of SimAir is to test policies at a strategic level, rather than function reactively at the operational time scale and incorporate all of the latest information that is available.

Two approaches for the SARP involving simulation were proposed by Guimarans et al. (2015), who used a Constraint Programming (CP) formulation combined with a Large Neighbourhood Search (LNS) algorithm to search the solution space. One algorithm used a deterministic objective function at each step of the heuristic until the search stopped. The resulting solution would then be simulated under a set of scenarios to test its robustness; if the robustness of the proposed solution is not considered satisfactory, its objective function value is added as a lower bound to the CP and the process repeats. The second approach uses the simulation at each step of the LNS search, called the SimLNS. Acceptance of the solution is either based on the average performance across the set or the worst performing replication. Whilst the variability was limited to just the flight times in the experimentation, the extension of this could be easily implemented, though this could harm the SimLNS computational performance.

Symbiotic simulation was defined at the Dagstuhl seminar on Grand Challenges for Modelling and Simulation (Fujimoto et al. 2002). The simulation receives measurements and observations from the physical system to improve its representation. The physical system can use the information gained from the what-if analysis of the simulation to improve its performance.

Aydt et al. (2009) contains a review of some of the important aspects of symbiotic simulation, from the purpose (such as whether it acts as a control system or a decision support system) to the triggering mechanism (reactive, preventative or pro-active). There have been many studies of symbiotic simulation since then with applications to various settings such as UAV collision avoidance (Holt et al. 2014) and quality control in gas turbine manufacturing (Meng et al. 2013).

Symbiotic simulation has been used to schedule tool operations in a semiconductor manufacturing setting by Aydt et al. (2011). This acts as a control system triggered pro-actively every 12 hours to find a new schedule for the next 24 hours. The schedule is found using an evolutionary algorithm.

Other applications include those to transportation settings. Vu et al. (2013) proposed a system that, given a traffic incident, could help an individual find a new route to their destination. Based on the predicted congestion from current data and a mesoscopic simulator, a finite set of routes is proposed and then simulated to find which has the smallest average travel time. Extending this idea to a population of vehicles, Aydt et al. (2012) focus on paths suggested by individual drivers to reduce the solution space whilst allowing a more holistic approach to traffic management within a city. This requires a great deal of information from individual vehicles to produce a high fidelity model.

Multi-fidelity modelling uses a simple approximation to the problem to identify candidate ‘best’ solutions which can then be simulated more thoroughly. These are also prevalent in simheuristics which often use a deterministic model to search the space and pick out promising solutions. One such example is the algorithm for combinatorial optimisation problems proposed by Juan et al. (2015). However, Xu et al. (2016) point out that the bias of the low-fidelity model is unknown and so could be a poor guide and criticise the naive approach of only simulating the solutions considered to be the best. Their proposed method (MO$^2$TOS) maps the solution space onto a one dimensional space using the low-fidelity model
rankings. The high-fidelity simulation is then allocated to solutions based on their ranking. This allows for a more thorough exploration during the simulation phase whilst still using the simulation selectively. Unfortunately, their work focusses on a setting in which all options can be evaluated with the low-fidelity model and there is no noise in the high-fidelity model output. Neither of these assumptions are applicable here.

3 MULTI-FIDELITY MODELS

A high-fidelity simulation model allows the user to be more confident that the simulation output is relevant to their concerns. Typically, however, the higher the fidelity, the more time is required to run the model. If the solution space is large and time is short, only a few solutions can be simulated. This is the case in disruption management. To use the simulation selectively, we propose using a simplified, deterministic model to help guide the search. This low-fidelity model takes the form of an integer programme. In this section we give an overview of the integer programme and simulation models followed by a description of how they are combined.

3.1 Low-fidelity Integer Programme

The use of an integer programme allows a straightforward way to represent some of the complexity of the airlines industry as well as a framework for solution methods. Our initial approximation is to use a deterministic model, neglecting potential recourse actions. This is a further simplification, though the stochastic elements are dealt with using the simulation.

The integer programme aims to allocate the aircraft to flights at minimum cost. It is based upon a time-space network (Zhang et al. 2015). Each node represents the time of a potential arrival or departure at an airport during the recovery period. Arcs either represent time spent at an airport (ground arcs) or potential rescheduled flights (flight delay arcs). Each flight has a series of delay options spaced equally between no delay and a maximum permissible delay. The allowed schedule changes are delaying flights, cancelling flights and exchanging aircraft assignments. Other options such as ferrying empty aircraft have not been considered. A feasible solution is a path for each aircraft through this time-space network using a sequence of ground and flight delay arcs.

The model includes a number of constraints. To respect minimum turn-around times of aircraft, only ground arcs representing sufficient time are generated. To ensure that the schedule can be operated beyond the recovery period, there must be sufficient aircraft at each airport after the recovery period (aircraft balance). Furthermore, airport runways have capacity constraints that must be respected, so that the airline can only use the runways during certain slots. It is assumed that this information is available to the airline. Other constraints include aircraft flow through the network and flight allocation constraints.

To optimise the integer programme, the Gurobi Optimizer 7.0.2 (Gurobi Optimization, Inc., 2017) is used. Preliminary results suggest that there is a trade-off between solution time and solution quality; see Figure 1. Increasing the number of aircraft involved in the problem or decreasing the delay spacing increases the time of the optimisation process. However, these increases in problem size also lead to lower costs as more rescheduling options become available.

3.2 High-fidelity Simulation Model

The simulation model consists of a single sub-fleet of an airline operating between a set of airports. Each aircraft in the sub-fleet aims to follow a schedule that is predefined. However, this is subject to variation in the flight times (due to the probabilistic flight duration and queueing times at airports), turn-around times and potential maintenance issues that need resolving. The schedule is stored within a database, which contains the origin, destination, departure and arrival times and the aircraft allocated for each flight. The schedules are the decision variables to be tested within the symbiotic simulation.
The sub-fleet population for the airline is created from a database and includes their identification, current maintenance state, number of flying hours since last maintenance, current location and when the next maintenance check is scheduled.

Much of the model structure is influenced by the SimAir design as presented in Lee et al. (2003). Unlike SimAir, however, crews and passengers are not considered. The simulation is implemented using AnyLogic 8.1.0 (The AnyLogic Company, 2017).

The data used by the simulation comes from https://www.flightradar24.com using the pyflightdata Python package (https://github.com/supercoderz/pyflightdata last accessed 4 October 2017). This data is currently used for two purposes: for inference into the flight duration distributions and the arrivals and departures at the airports relevant to the airline.

Each flight in the model consists of a series of events, as shown in Figure 2. The core parts of the model are the flight times, the take-off and landing queues and the time at the airport. Each of these are sources of stochasticity within the system and so are important to consider.

The flight times are sampled from a log-logistic distribution with Maximum Likelihood Estimates for the parameters based on the data.

The landing and take-off queues are modelled as $G(t)/D(t)/1$ queueing systems. Each airport has a schedule of arrivals and departures, which is assumed known to the airline. However, the aircraft rarely arrive at the queues on time and so some deviation is added in the simulation. The distribution of deviations
for airport arrivals is normal and is log-logistic for departure, both are estimated from the data. The deterministic service time at the queues represents the minimum spacing required between aircraft in the procedures. It is assumed that, given the time of day and weather conditions, this is a deterministic function.

An aircraft’s length of stay at an airport is governed by whether or not the aircraft requires maintenance. These processes are not tracked in our current data source and so some assumptions have been made about the lengths of time involved and the distributions. The exact processes required in the turn-around times are not modelled explicitly. Turn-around times are assumed to be normally distributed. Planned maintenance is assumed to have a fixed duration whilst unplanned maintenance is sampled from a Gamma distribution. The parameters for these distributions vary depending on the services available at individual airports. The aircraft will either leave when scheduled or once they are ready, which ever is later.

The time to failure model used is a Weibull distribution conditional on the time spent flying since the aircraft’s last maintenance event. This assumes that the aircraft is a single component and so a more sophisticated model would be required in practise. An aircraft fails once its time in flight exceeds its failure time.

3.3 Combining the Models

The low-fidelity model is used to generate a set of potentially good solutions. This process dramatically reduces the solution space for the simulation optimisation procedure. We are currently considering two different algorithms for generating this set. The first is to use the Gurobi Optimizer 7.0.2 to find the set of $M$ best solutions to the problem. Firstly it finds the optimal solution before searching other branches for the next best.

The second is to use the $\varepsilon$—Constraint method with secondary objectives such as total delay, number of cancellations and the number of aircraft swaps made. By varying the limits of these constraints and solving the problem multiple times, a variety of solutions on the efficient frontier can then be found.

Part of our investigation is to look at which of these methods perform better and if there are particular circumstances under which one outperforms the other.

Once this set has been found, a Ranking and Selection procedure can be implemented. Whether or not screening or a sequential methodology is selected would depend upon the size of the set from the integer programme and the time still available to the operations control centre of the airline.

An extension of this approach would be to use the simulation to perform a local search around each of the solutions provided by the low-fidelity model. This would allow a more thorough search of the solution space guided by the low-fidelity model, and may help understand inaccuracies in the low-fidelity model. However, the definition of a neighbourhood in a space of scheduling is non-trivial due to the strong dependence between all variables. Making adjustments to one variable inevitably leads to (potentially many) more adjustments being required. Part of our future work would be to find the most appropriate dimensions in which the neighbourhood could be defined.

4 DEMONSTRATION

In this section, we will discuss some preliminary results of the combination of the integer programme and the simulation model. Here we use a small problem as an indication of the intended use and the potential results. The example problem involves a fleet of five aircraft operating between London Heathrow, Edinburgh and Manchester airports, with 83 flights between them over two days. One of the aircraft has been found to have a mechanical failure and is expected to be unavailable until 1 hour and 20 minutes after it was scheduled to fly. The aim is to find a revised schedule for the next two days.

The costs are assumed to only arise from passenger compensation and delays (delay cost is assumed to be proportional to the delay length). The only slot constraints considered are for overnight arrivals and departures. The maximum allowable delay is 3 hours with a delay interval of 10 minutes. To allow for some slack, the low-fidelity model assumes a minimum turn-around time of 50 minutes, whilst the mean
turn-around time in the simulation is 40 minutes, with a standard deviation of 10 minutes. The resulting integer programme has 4748003 variables, 164834 constraints and more than 2450 feasible solutions.

The Gurobi Optimizer was used to return the best five solutions to the integer programme found within a time limit. These solutions were then compared in the simulation using the Ranking and Selection procedure proposed by Rinott (1978). The desired probability of correct selection was 0.95 with an indifference-zone parameter of $d = £1000$. The costs estimated by both the integer programme and the simulation are shown in Figure 3.

![Figure 3: The estimated expected costs from both models. The simulation include 95% confidence intervals for the mean cost.](image)

The plot suggests that the deterministic model significantly underestimates the costs of operating the proposed schedule changes in reality. This is to be expected as adding variability into queueing systems tends to increase the time in the system. However, the ranking of the solutions by each model is similar, which is an important aspect of multi-fidelity modelling. This is encouraging as it suggests that a good solution in the low-fidelity model can translate into a good solution for the high-fidelity model.

5 DISCUSSION AND FUTURE WORK

The results of our initial experimentation have been promising. They indicate that there may indeed be a correlation between the performance estimates of the two models. However, there are a number of issues that are still to be resolved.

Several questions remain regarding the use of the low-fidelity model, particularly the search for solutions. For example, how many solutions should be looked for? Would a more heuristic approach that produced more ‘good’, defined using an optimality gap, rather than optimal solutions be more appropriate? There are also parameters within the search that could be tuned. Further experiments on larger examples are required and may help to determine which of the methods outlined in Section 3.3 are more appropriate for this application.

The main aim of multi-fidelity modelling is to use information from a simple approximation to the problem to influence how the high-fidelity model is used. Therefore, the more representative that approximation is, the better the information is and so the better the guidance is. Our low-fidelity model
is completely deterministic, which is a very strong assumption. If it could be replaced by a model that incorporated some stochastic elements, the gap between the two models should decrease which should in turn improve the solutions suggested. However, the computational efficiency of the low-fidelity model must be preserved. In practise, such a tool would be used repeatedly. Thus, one could ‘learn’ how well the low-fidelity model ranks the solutions compared to the simulation model. Therefore, one could have a calibration process in which a set of parameters of the low-fidelity model are gradually tuned to improve the information it provides.

Whilst our current method does take into account the inaccurate performance estimation of the low-fidelity model within the group of proposed solutions, it assumes that the inaccuracy does not vary significantly, and behaves systematically across the whole space. That is, it assumes that no solution’s rank differs substantially between the two models. However, the inaccuracy may not behave in a predictable manner. Thus, there may be solutions that were not selected by the low-fidelity model that have greater performance than those that are. This motivates a greater involvement of the simulation in the process. This could be using a local neighbourhood search around the solutions in the proposed set. As part of this development, a new formulation of the solution space would be required in order to define the neighbourhoods to search through. An alternative would be to use a more iterative approach between the high and low-fidelity models such as that considered in other simheuristics.

The ideal use of this procedure would be to offer the airlines evidence on the performance of a small set of revised schedules across a number of performance measures. For example, an airline may be willing to increase the expected cost if this came with a substantial decrease in the variance. A lower variance may help to reduce the risk of very high costs, which is an important factor airlines would want to consider.

There are also developments that could be made to the simulation model. For example, a more sophisticated model of the flight durations to include weather conditions and deliberate changes in speed made by the pilot. This could also be incorporated into the solution space of the problem.

In conclusion, we have proposed a multi-fidelity modelling approach to airline disruption management, combining an integer programme with a simulation model. The early results suggest that this is a promising method, though more development and experimentation is still required.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support of the EPSRC funded EP/L015692/1 STOR-i Centre for Doctoral Training and the NSF Grant CMMI-1068473. We would also like to thank Nigel Jackson, Richard Standing and Mike Chester for the original research idea and contextual information.

REFERENCES


AUTHOR BIOGRAPHIES

LUKE RHODES-LEADER is a Ph.D. student of the Statistics and Operational Research Centre for Doctoral Training in Partnership with Industry at Lancaster University.

BHAKTI STEPHAN ONGGO is an Associate Professor of Data Analytics at Trinity Business School, Trinity College Dublin, The University of Dublin, Ireland. His research interests lie in the areas of predictive analytics using simulation (data-driven simulation, hybrid modelling, agent-based simulation, discrete-event simulation) with applications in operations and supply chain management (e.g. hospital, manufacturing, transportation) and social science (e.g. spread of perception). He is the associate editor for the Journal of
DAVID J. WORTHINGTON is a senior lecturer in Operational Research in the Department of Management Science in Lancaster University Management School. He researches the modelling and management of time-dependent queueing systems, and applications of Management Science in healthcare. These research interests often coincide.

BARRY L. NELSON is the Walter P. Murphy Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University and a Distinguished Visiting Scholar in the Lancaster University Management School. He is a Fellow of INFORMS and IIE. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems, and he is the author of *Foundations and Methods of Stochastic Simulation: A First Course*, from Springer.