

MEASURING THE ROBUSTNESS OF AIRLINE FLEET SCHEDULES

F. Bian¹, E. K. Burke², S. Jain³, G. Kendall², G. M. Koole⁴, J. D. Landa Silva², J. Mulder⁵, M. C. E. Paelinck⁵, C. Reeves⁶, I. Rusdi⁷, M. O. Suleman¹

¹*University of Oxford, UK*

²*The University of Nottingham, UK*

³*Aston University, UK*

⁴*Free University of Amsterdam, The Netherlands*

⁵*KLM Airlines, The Netherlands*

⁶*Coventry University, UK*

⁷*Technical University of Delft, The Netherlands*

Abstract Constructing good quality fleet schedules is essential for an airline to operate in an effective and efficient way in order to accomplish high levels of consumer satisfaction and to maximise profits. The robustness of an airline schedule is an indicative measure of how good the schedule is because a robust plan allows the airline to cope with the unexpected disturbances which normally occur on a daily basis. This paper describes a method to measure the robustness of schedules for aircraft fleet scheduling within KLM Airlines. The method is based on the "Aircraft on Ground (ACOG)" measure, it employs statistical methods (although alternative methods were also considered) and it is shown to provide a good estimation of the robustness of a given schedule.

Keywords: modelling, airline scheduling, schedule quality measures.

1. INTRODUCTION

The problem of generating fleet schedules is crucially important to the efficiency of an airline (Barnhart *et al.*, 1997; Barnhart and Talluri, 1997). An effective schedule can lead to significant savings. It can also, and perhaps more importantly, contribute to higher levels of customer satisfaction. Customers who experience regular delays with a particular airline are likely to take their custom elsewhere. Of course, delays are inevitable for a wide range of reasons (e.g. technical breakdowns, security alerts, adverse weather, etc). However, an indicative measure of the quality of an airline schedule is its level of *robustness*: How well can a schedule cope with a delay(s) to a particular aircraft(s)?

Is there enough slack in the schedule to minimise the knock on effect of a delay to a particular aircraft? If there is no slack in the schedule then a delay to one aircraft could affect a significant proportion of the fleet and this could have major resource implications. If passengers miss connecting flights then the airline has to cover the incurred costs. However, building slack into the schedule is expensive. It essentially involves aircraft standing idle. One of the goals in trying to generate a high quality fleet schedule is to build in enough slack to ensure that the schedule has an acceptable level of robustness while, at the same time, attempting to keep costs at an effective level. It would be very easy indeed to build a very robust schedule. However, it would be too expensive to implement. It would also be possible to build a schedule which minimises cost by decreasing aircraft idle time. However, this could easily lead to an increase in the overall incurred costs if one minor delay to one aircraft leads to a chain of delays. In summary, the goal is to provide an effective balance between robustness and aircraft idle time.

The integration of schedule optimisation algorithms and other systems in an airline is crucial to achieve an effective scheduling environment that considers all functions of the airline (Mathaisel, 1997). Reviews of research on airline scheduling are presented in Etschmaier and Mathaisel (1985) and Richter (1989). A more recent survey on models and solution methods for a range of problems in aircraft scheduling was carried out by Gopalan and Talluri (1998).

Aircraft scheduling is often addressed simultaneously with other associated problems. An example is provided by fleet assignment with time windows where the assignment of aircraft is carried out simultaneously to scheduling flight departures in order to improve flight connection opportunities and minimise costs (Rexing *et al.*, 2000). The scheduling of maintenance operations and of aircraft are considered simultaneously using network models and a two-phase heuristic by Feo and Bard (1989), while crew availability and maintenance operations are taken into account while tackling the fleet assignment problem in Clarke and Hane (2001). The additional constraint of equal aircraft utilisation when tackling fleet assignment and aircraft routing problems is considered by (Barnhart *et al.* (1998). A network model for large-scale fleet assignment problems that permits the expression of constraints within a unified framework was presented by Rushmeier and Kontogiorgis (1997).

Integer linear programming techniques have been applied by several researchers to tackle fleet assignment, aircraft routing and related problems (Abara, 1989; Subramanian, 1994; Hane *et al.*, 1995). Dynamic programming and heuristics have also been investigated for the problem of fleet assignment (El Moudani and Mora-Camino, 2000). Recently, meta-heuristic methods have been used to tackle airline scheduling problems. For example, simulated annealing was applied to the optimisation of airline schedules by Mashford and Marksjo (2001). Sosnowska and Rolim (2001) showed that by applying sim-

ulated annealing to the fleet assignment and aircraft routing, improvements of about 10–20% over the method used by the company could be achieved. A genetic algorithm was applied to generate alternative routes for air traffic by Oussedik *et al.* (2000). Also recently, genetic search methods have been applied to solve the problem of sequencing the arrival of aircraft in airports (Hansen, 2004; Ciesielski and Scerri, 1997, 1998).

Re-scheduling is a crucial activity for airlines and it has to be carried out on a daily basis due to a number of uncertainties and unforeseen events. Disruptions of planned schedules can result in a chain of events that can cause major disruptions throughout the system. A survey of techniques employed to recover from these disruptions is presented by Filar *et al.* (2001). A stochastic model is employed by Rosenberger *et al.* (2003) to show that the actual performance of an airline differs greatly from the planned performance while Argüello and Bard (1997) propose a GRASP method to reconstruct schedules while minimising costs and satisfying constraints. Network models and Lagrangian relaxation were used by Yan and Lin (1997) for aircraft re-scheduling given a specific disruption that affects the airline operations greatly and causes substantial decrements in profits and levels of service: the temporary closure of airports (see also Thengvall *et al.*, 2001, 2004). The problem of changing the assigned aircraft to specific flights while satisfying existing constraints is addressed by Jarrah (2000), Talluri (1996) and Klincewicz and Rosenwein (1995). A steepest ascent local search heuristic was applied by Love *et al.* (2002) to re-schedule aircraft and it was capable of finding good quality schedules in a short amount of time.

The problem that is addressed in this paper is discussed in the next section. It represents a real-world problem that faces KLM Airlines on a daily basis.

2. PROBLEM DESCRIPTION

Within KLM, two departments are responsible for the fleet schedule. The network planning department produces schedules which are then passed to the operations department who has the responsibility for implementing them and running them on a day-to-day basis. These two departments have conflicting objectives. The network department aims to produce a schedule which is as cost effective as possible. This essentially means maximising aircraft usage by minimising their idle time. The operations department prefers schedules that have enough slack to ensure a certain level of robustness. This means having as much aircraft idle time as possible. Then, the overall aim is to produce a schedule with the right balance between these two conflicting objectives briefly described above.

The aim for KLM is to introduce a method that checks the robustness of a schedule, from the network department, before it is passed to the operations

department for implementation. One way to achieve this is to run a simulation. However, this is seen as too time consuming and other methods are sought to test for the robustness of the schedule.

KLM flies to over 150 destinations using 97 aircraft. Four times a year, a new flight schedule is developed. Though the operational feasibility is taken into account to a certain degree during the development process, the aim at that stage is largely to maximise the number of seats that can be sold. During schedule development, KLM considers various commercial aspects such as the expected demand per destination and the number of possible transfer connections at Schiphol Airport in Amsterdam.

The realisation of a flight schedule involves a number of parties. As described above, the initial plan is developed by KLM's network planning department. The initial plan is based on commercial and strategic insights and long term plans for the fleet composition, cabin crew and baggage handling.

Two months before the beginning of a schedule plan, the plan is handed over to the operational department, the Operation Control Centre. From that moment on they are the owners of the plan and small adaptations have to be evaluated and approved by them. This department will try to prevent and solve problems such as emergencies and bottlenecks and, in case of unsolved problems, try to minimize the effects on succeeding flights. A final plan is created two weeks before the beginning of the plan where passenger bookings are matched with aircraft capacities

In order to monitor the performance of a flight schedule, some critical performance indicators are defined. These are

- The departure and arrival punctuality: that is, the percentage of flights that departed or arrived on time.
- The completion factor: that is, the percentage of accomplished flights. These are all flights that were not cancelled.
- The No Connection Passenger factor: that is, the percentage of transfer passengers that missed their connections due to operational problems.
- The Irregularity-rate: that is, the number of bags that were not delivered on time.

For the punctuality performance indicator the contribution of each of the involved parties is also monitored. This introduces the concept of building blocks. The whole operational process is divided into sub processes, (the so called building blocks). Each building block is owned by a *capacity and service provider*, these being Ground Services, Front Office, Air Traffic Management, Engineering and Maintenance, Cabin and Cockpit Crew, Cargo and Operations Control. Seven Building Blocks have been established, these are called

- BB1: Flight
- BB2: Arriving aircraft
- BB3: Layover aircraft
- BB4: Departing aircraft
- BB5: BB5.1 Transferring passengers
BB5.2 Transferring baggage
- BB6: BB6.1 Arriving passengers
BB6.2 Arriving baggage
- BB7: BB7.1 Departing passengers
BB7.2 Departing baggage.

A diagrammatical representation of the temporal sequence of the building blocks and their relationships to each other is shown in Figure 1. These have been delimited in order to provide clear process distinction as well as accountability.

The doors being opened and closed are the points at which responsibility passes from one capacity and service provider to another. The distinction of the *first door* being opened is made because a door can either be the passenger door(s) or a baggage door(s). For example, once a plane has physically landed it is not actually considered to have *landed* (i.e. with responsibility passed to the ground staff) until one (passenger OR baggage) door has been opened. In contrast, responsibility changes back again when *all* doors have been closed, not just one door.

All these agreements and the flight schedule itself comes together into an operational plan. This functions as a contract between Network, Operations Control and the Building Blocks (Capacity & Service Providers). The plan covers an operational plan period of between 2 to 4 months spread over the year. It consists of agreements concerning a schedule plan and a capacity plan position for each specific period. It contains a demand-driven schedule that has been fully checked with the Building Block representatives (Capacity & Service Providers) and Operations Control by means of an operational check. Eventually the agreements enable each provider to deliver an operational performance forecast. This could deviate from the targets as laid down in the corresponding Business Plan. Each operational plan will be finalized two months (at the latest) prior to each operational plan period.

The schedule is usually published as an Aircraft Rotation Schedule, which is different each week. This is due to the fact that each day many adaptations are made so as to minimise delays. For instance, if KLM know that an aircraft will arrive at Schiphol Airport with a delay, they could assign its next flight to

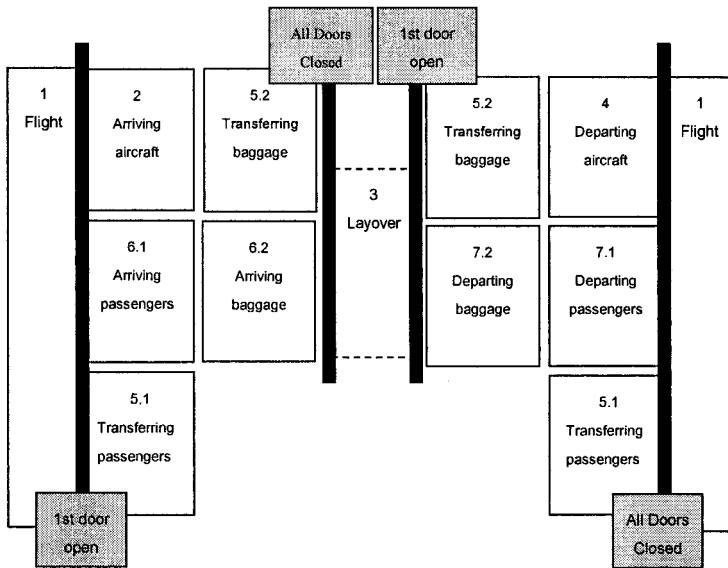


Figure 1. Figure 1. Building blocks sequence and relationships.

another aircraft so that that flight can still leave on time. Usually, KLM will also need other adaptations to have all flights fit into the Rotation Schedule again. When a schedule is first published, KLM do not know the exact layout of the Rotation Schedule, so they publish a hypothetical “average” one instead.

Before a schedule is published, an estimation of the expected punctuality (that is the percentage of “on time” flights) is performed using a simple deterministic model. As this model lacks accuracy, a simulation model is currently being developed in order to enable a better forecast. This model simulates aircraft movements according to a given schedule. The model subjects the schedule to a “stress test” by generating various disruptions such as air traffic congestion, delays during the boarding process or unexpected problems during maintenance. Throughout the simulation, a *Problem Solver* algorithm attempts to resolve delays by swapping flights in the Rotation Schedule, or in extreme cases, by cancelling flights. More successful runs of the simulation are considered as better schedules for implementation.

A simulation, though, has several disadvantages. Processing times are usually too long, which limits the number of schedules that can be assessed. Also, KLM need to collect a huge amount of data about the processes that are being simulated. For the simulation model currently under development they need statistics about the variation in the actual flight duration, the variation in the

time it takes to handle an aircraft on the ground (boarding, fuelling, catering, etc), breakdown times of each aircraft type, etc. Each of these statistics must constantly be updated to reflect the change in flight routes, working methods, fleet, etc.

KLM are currently seeking a simpler model that would enable them to make a comparative statement, such as “Of a number of alternative schedules, schedule X will provide the best performance.”

3. MODELS FOR THE PROBLEM

It was anticipated that there should be some features of any schedule that would be correlated with its performance. The first question is then what features should be investigated? A brainstorming session with representatives of KLM led to some suggestions. It was expected that the number of potential swaps available to a delayed flight would be an important factor, but measuring this value was not easy. In practice, it might also be necessary to undertake a cascade of swaps, so another possible measure of performance would be the length of time and/or the number of swaps needed to restore the schedule to its normal condition. However, this is also complicated to determine, although the *Problem Solver* module of the simulation could be invoked if necessary.

After further discussion, it was agreed to look at a simpler measure, which could easily be found, and is arguably a surrogate for some of the more complex measures suggested. This is the “Aircraft on Ground” (ACOG) measure which gives an indication of the number of aircraft on ground. ACOG can be calculated from the number of arriving aircraft, the number of layover aircraft, and the number of departing aircraft. Having obtained some features related to this measure, the next step is the identification of a suitable model for purposes of prediction. Candidates here include multiple linear regression methods, regression trees, neural nets and other pattern recognition techniques. However, the fact that the amount of data available was small meant that data-hungry methods should be avoided if at all possible. Thus it was resolved to begin the investigation with traditional statistical methods.

4. EXPERIMENTAL RESULTS

Eleven schedules were available (Summer/Winter 2000–02, apart from the last 13 weeks of 2002). KLM’s operation at Schiphol is such that the activity occurs in four major waves—a deliberate strategy to maximise passengers’ opportunities for making onward connections. Graphing the number of aircraft available on the ground reveals this pattern clearly. These can be counted in two ways: the more accurate picture is obtained by subtracting the lengths of BB2 and BB4, leaving just those aircraft that are actually idle at a given moment.

Table 1. Performance indicators for departure punctuality and arrival punctuality using two different models. For the predictor sets: p1, 1st peak; m, first moment (mean); sd, second moment (standard deviation); sk, third moment (skewness); k, fourth moment (kurtosis).

	Using BB3 only	Using BB2–4
PI Departures		
Predictor sets	p4m, p1sd, p1sk, p1k	p2m, p4m, p2sd, p4sd, p3sk
<i>R</i> -squared	95.6%	91.6%
<i>p</i> -value (F-test)	0.00032	0.01028
PI Arrivals		
Predictor sets	p4m, p1sk, p3sk, p3k	p1m, p4m
<i>R</i> -squared	95.2%	84.1%
<i>p</i> -value (F-test)	0.00042	0.00064

However, it is a simpler calculation to count the whole of the time on the ground from “First Door Open” to “Last Door Closed”, which comprises the whole of BBs 2,3 and 4.

In the case of European operations, each day is more or less identical, so peaks can be defined quite easily. For each peak, the first four moments of the ACOG values were calculated for each day, using both definitions—BB3 and BB234. As days are so alike (apart from the very first day of a new schedule), one day can be selected at random as a representative of a schedule. As there are four peaks daily, we have 16 features as inputs, which we need to associate with the performance indicators (PIs) already calculated by KLM. The ones used for the models developed here were simply the departure and arrival punctualities: the fraction of planes (of those scheduled) that departed or arrived on time.

As a first step, correlations were calculated between the PIs and the 16 input variables. The six or seven most highly correlated input variables were then used in a stepwise regression procedure (using S-plus) to determine the best balance between parsimony and explanatory power (S-plus uses the Akaike information criterion for this purpose.) Table 1 summarises the models determined by this approach.

Of interest is the fact that p4m, the mean number of ACOG, is important for all four models, but the other predictors seem to be far less important. From KLM’s point of view, this does not matter if the predictions are good enough, but from a modeller’s perspective we would like to see more consistency. However, all models are based on just 11 data points, so perhaps the lack of consistency is not surprising. Prediction intervals can easily be obtained on the assumption of Normally distributed errors: these vary from $\pm 2\%$ for punctualities in the middle of the range to $\pm 3\%$ at the edges.

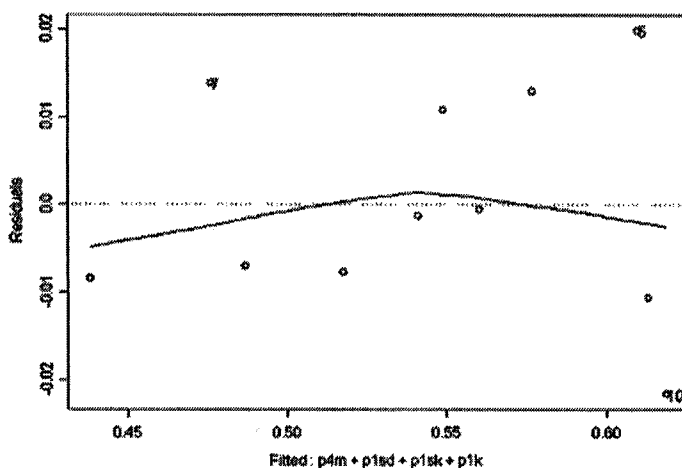


Figure 2. Residuals against fitted values for departure punctuality using BB3 only.

It was surprising that the R -squared values were as high as they were—we were anticipating that a linear model would be too simple, yet it seems quite powerful. Of course regression analysis makes certain assumptions about the errors, and it is necessary to check the residuals to see if these assumptions are plausible. The plot of residuals against fitted values was obtained for each model; in no case does a systematic pattern seem plausible, and a random scatter is obtained, as shown in Figure 2.

The three most extreme outliers (points 5, 7 and 10) are labelled; point 5 might well have been affected by September 11, but possible reasons for the others are not known. A smooth has been applied, but its slopes are not very steep, so the assumption that the errors are independent random variables seems plausible. Similar graphs were obtained for the other three models.

QQ plots of the residuals against Normal quantiles were also obtained. Figure 3 below shows the same case as in Figure 2.

The tails of the distribution in particular are not well fitted, so the assumption that the errors are Normally distributed is perhaps questionable. Thus any confidence intervals should be treated cautiously. In any case, the response variable in all four models is actually a ratio that is confined to remain between 0 and 1. This means that a better theoretical model would be based on a logistic transformation, since it is theoretically possible that a simple linear model could generate predictions outside the possible range of values. For example, we can hardly have a punctuality of greater than 100%! Such a model would also be based on a more plausible probability model than the Normal distribution.

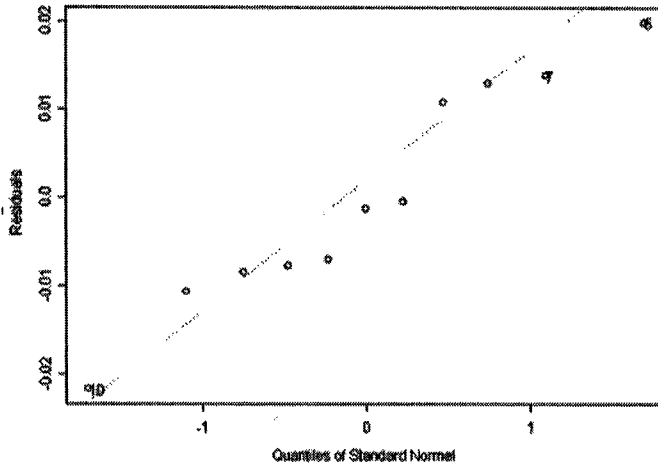


Figure 3. Normal QQ plot for residuals for departure punctuality using BB3 only.

However, attempts to fit such a model did not produce an improvement. A possible explanation is that the data available are all in the region of approximate linearity of the logistic curve. Consequently, any attempt to identify the turning points of the curve is likely to be rather speculative. In any case, on inspecting the coefficients of the models, it seems unlikely that we would predict bizarre fractions in practice. For example, using the most extreme values observed in the first model above would predict only 80% departure punctuality, and in the opinion of KLM's experts it is hard to imagine physical circumstances in which these values could be exceeded simultaneously (there is just not enough space to put many more planes, for example).

Thus, despite the attractions of a more plausible theoretical model, the airline is comfortable with the predictive ability of a simpler linear model.

5. CONCLUSIONS

An analysis of the expected number of aircraft on the ground has been shown to provide a good prediction for the robustness of a given schedule. Further refinements are possible—and desirable—but even this work has given the KLM's operations department a better insight into what makes a fleet schedule easier or harder to implement effectively. Some of the work that still needs to be done includes an analysis of the effect of day-to-day variations in the schedule—these variations are small, but preliminary work has suggested that the definition of activity peaks needs to be tighter, and the possibility of a day-of-the-week effect should also be explored. Furthermore, the schedules

examined so far have concentrated only on the European operations, where fleet homogeneity is substantial and diurnal variation is small. Incorporating the effects of the inter-continental timetable may lead to some changes in these conclusions.

Acknowledgments

We would like to thank the following bodies for supporting the collaboration described in this paper:

- The Lorentz Centre in Leiden: 45th European Study Group for Mathematics with Industry.
- The EU Mathematics, Computing and Simulation for Industry Network (MACSI-net).
- The Smith Institute for Industrial Mathematics and Systems Engineering.

References

- Abara, J. (1989), Applying integer linear programming to the fleet assignment problem. *Interfaces*, **19**:20–28.
- Argüello, M. F. and Bard, J. F. (1997), A GRASP for aircraft routing in response to groundings and delays. *Journal of Combinatorial Optimization*, **1**:211–228.
- Barnhart C., Boland N. L., Clarke L. W., Johnson E. L., Nemhauser G. L. and Shenoi, R. G. (1998), Flight string models for aircraft fleet and routing, *Transportation Science*, Special Issue on Airline Optimization, **32**:208–220.
- Barnhart C., Lu, F. and Shenoi, R. (1997), Integrated airline scheduling. In *Operations Research in the Air Industry*, Gang Yu (Ed.), International Series in Operations Research and Management Science, Vol. 9, Kluwer, Dordrecht, pp. 384–403.
- Barnhart, C. and Talluri, K. (1997), Airline operations research. In *Design and Operations of Civil and Environmental Engineering Systems*, A. McGarity and C. ReVelle (Eds.), Wiley, New York, pp. 435–469.
- Ciesielski V. and Scerri P. (1997), An anytime algorithm for scheduling of aircraft landing times using genetic algorithms, *Australian Journal of Intelligent Information Processing Systems*, **4**:206–213.
- Ciesielski V. and Scerri P. (1998), Real time genetic scheduling of aircraft landing times. In *Proceedings of the 1998 IEEE International conference on Evolutionary Computation (ICEC98)*, IEEE, Piscataway, NJ, pp. 360–364.
- Clarke L. W., Hane C. A., Johnson E. L. and Namhauser G. L. (2001), Maintenance and crew considerations in fleet assignment, *Transportation Science*, **3**:249–260.
- El Moudani, W. and Mora-Camino, F. (2000) A dynamic approach for aircraft assignment and maintenance scheduling by airlines, *Journal of Air Transport Management*, **6**:233–237.
- Etschmaier, M. and Mathaisel, D. (1985) Airline scheduling: an overview, *Transportation Science*, **19**:127–138.
- Feo, T. A. and Bard, J. F. (1989) Flight scheduling and maintenance base planning, *Management Science*, **35**:1415–1432.

- Filar, J. A., Manyem, P. and White, K. (2001) How airlines and airports recover from schedule perturbations: a survey, *Annals of Operations Research*, **108**:315–333.
- Gopalan, R. and Talluri, K. T. (1998) Mathematical models in airline schedule planning: a survey, *Annals of Operations Research*, **76**:155–185.
- Hane, C. A., Barnhart, C., Johnson, E. L., Marsten, R. E., Nemhauser, G. L. and Sigismondi, G. (1995) The fleet assignment problem: solving a large-scale integer program, *Mathematical Programming*, **70**:211–232.
- Hansen, J. V. (2004) Genetic search methods in air traffic control, *Computers and Operations Research*, **31**:445–459.
- Jarrah, A. I. (2000) An efficient airline re-fleeting model for the incremental modification of planned fleet assignments, *Transportation Science*, **34**:349–363.
- Klincewicz, J. G., Rosenwein, M. B. (1995) The airline exception scheduling problem, *Transportation Science*, **29**:4–16.
- Love, M., Sorensen, K. R., Larsen, J. and Clausen, J. (2002) Disruption management for an airline—rescheduling of aircraft, *Applications of Evolutionary Computation: Proceedings of the EvoWorkshops 2002*, Lecture Notes in Computer Science, Vol. 2279, Springer, Berlin, pp. 315–324.
- Mashford, J. S, Marksjo, B. S. (2001) Airline base schedule optimisation by flight network annealing, *Annals of Operations Research*, **108**:293–313.
- Mathaisel, D. F. X. (1997) Decision support for airline schedule planning, *Journal of Combinatorial Optimization*, **1**:251–275.
- Oussedik, S., Delahaye, D. and Schoenauer, M. (2000) Flights alternative routes generator by genetic algorithms. In *Proceedings of the 2000 Congress on Evolutionary Computation (CEC 2000)*, IEEE, Piscataway, NJ, pp. 896–901.
- Rexing, B., Barnhart, C, Kniker, T., Jarrah, A. and Krishnamurthy, N. (2000) Airline fleet assignment with time windows, *Transportation Science*, **34**:1–20.
- Richter, H. (1989) Thirty years of airline operations research, *Interfaces*, **19**:3–9.
- Rosenberger, J. M., Schaefer, A. J., Goldsman, D., Johnson, E. L., Kleywegt, A. J. and Nemhauser, G. L. (2003) A stochastic model of airline operations, *Transportation Science*, **36**:4.
- Ruland, K. S. (1999) A model for aeromedical routing and scheduling, *International Transactions in Operational Research*, **6**:57–73.
- Rushmeier, R. A. and Kontogiorgis, S. A. (1997) Advances in the optimization of airline fleet assignment, *Transportation Science*, **31**:159–169.
- Sosnowska, D. and Rolim, J. (2001) Fleet scheduling optimization: a simulated annealing approach. In *Practice and Theory of Automated Timetabling III (PATAT 2000)*, E. Burke and W. Erben (Eds.), Lecture Notes in Computer Science, Vol. 2079, Springer, Berlin, pp. 227–241.
- Subramanian, R. (1994) Coldstart: fleet assignment at Delta Air Lines, *Interfaces*, **24**:104–120.
- Talluri, K. T. (1996) Swapping applications in a daily airline fleet assignment, *Transportation Science*, **30**:237–248.
- Thengvall, B. G., Yu, G. and Bard, J. F. (2001) Multiple fleet aircraft schedule recovery following hub closures, *Transportation Research A*, **35**:289–308.
- Thengvall, B. G., Bard, J. F. and Yu, G. (2004) A bundle algorithm approach for the aircraft schedule recovery problem during hub closures, *Transportation Science*, to appear.
- Yan, S. and Lin, C. G. (1997) Airline scheduling for the temporary closure of airports, *Transportation Science*, **31**:72–82.