

A three-stage approach for aircraft line maintenance personnel rostering using MIP, discrete event simulation and DEA

***Jorne Van den Bergh
Philippe De Bruecker
Jeroen Beliën
Liesje De Boeck
Erik Demeulemeester***

HUB RESEARCH PAPERS 2012/42
ECONOMICS & MANAGEMENT
NOVEMBER 2012

A three-stage approach for aircraft line maintenance personnel rostering using MIP, discrete event simulation and DEA

Jorne Van den Bergh^{a,c,†}, Philippe De Bruecker^{b,d}, Jeroen Beliën^{a,c}, Liesje De Boeck^{a,c} and
Erik Demeulemeester^b

^a HUBrussel, Center for Informatics, Modeling and Simulation, Warmoesberg 26, 1000 Brussels
(Belgium), jorne.vandenbergh@hubrussel.be, jeroen.belien@hubrussel.be,
liesje.deboeck@hubrussel.be

^b KU Leuven, Research Center for Operations Management, Naamsestraat 69, 3000 Leuven
(Belgium), philippe.debruecker@kuleuven.be, erik.demeulemeester@kuleuven.be

^c Affiliated researcher KU Leuven, Research Center for Operations Management, Naamsestraat 69,
Leuven (Belgium)

^d Affiliated researcher HUBrussel, Center for Informatics, Modeling and Simulation, Warmoesberg
26, 1000 Brussels (Belgium)

[†] Corresponding author at: Center for Informatics, Modeling and Simulation, Warmoesberg 26, 1000
Brussels (Belgium). Tel: +32 2-210 16 11; fax: +32 2 217 64 64.

Keywords: line maintenance, staff allocation, uncertainty, scheduling, tabu search, Data Envelopment
Analysis

Abstract

Personnel scheduling problems need to cope with personnel preferences, coverage constraints, legal restrictions, and many other constraints. We present a three-stage methodology that can be used to select personnel rosters. In the first stage we generate multiple personnel rosters with a mathematical programming model. In the second stage, the performance of the rosters regarding a number of service criteria is evaluated through discrete event simulation. In the third stage, a ranking is made using Data Envelopment Analysis. The methodology is tested on a personnel scheduling problem for aircraft line maintenance.

1 Introduction

Line maintenance includes regular short inspections of an aircraft between arrival and departure at an airport (Gupta, Bazargan, & McGrath, 2003). Once a flight schedule is established, the maintenance company can assign a maintenance schedule to each maintenance station. Based on this schedule, the company builds a staffing model, considering the fleet type, number and type of maintenance jobs, specific client requests, etc. Efficient personnel scheduling not only can reduce costs, but is also directly related to flight safety issues (Yan, Yang, & Chen, 2004). Maintenance workforce schedules have to meet all safety requirements and time constraints: if timetable punctuality is affected, flight delays as well as extra operating costs are incurred. According to delay analyses carried out by Eurocontrol, around 50% of delays are due to airline operational processes, while the remaining delays are due to other causes such as air traffic control, weather and airport capacity constraints (Eurocontrol, 2011). The proportion of intra-European flights delayed by more than 15 minutes has decreased in the past few years (CODA, 2010). Nevertheless, it was still larger than 15% in 2009 (CODA, 2010). Those delays can have a major impact on airline operations in general and line maintenance operations in particular. Late maintenance that causes primary delays implies causality delays at other airports and huge costs as a consequence. The workforce schedules should be able to prevent this.

Real-world personnel scheduling problems are typically over-constrained. One has to consider coverage constraints, personnel preferences, workload balancing, legal restrictions, etc. Most papers try to mimic this behavior with multi-objective programming. A mathematical programming model is used to generate an optimal roster regarding these constraints and (weighted) objectives. In this paper, we present an alternative way of tackling personnel scheduling problems by using mathematical programming, discrete event simulation and Data Envelopment Analysis (DEA). With this approach, the company is supplied with detailed information about the rosters, which they can use to make the selection objectively. First, a mathematical model generates multiple feasible low cost rosters, which are evaluated with simulation. The roster characteristics and the simulation results are then fed into a DEA model which then identifies the efficient rosters.

This methodology is applied to a case study, in which we evaluate airline maintenance workforce rosters in the occurrence of flight delays (i.e., when flights depart after their scheduled time of departure (STD)). These rosters are the results of the mixed integer linear programming (MILP) model of Beliën, Demeulemeester and Cardoen (2012). In their model, they try to protect the maintenance schedule from flight arrival delays by scheduling no maintenance jobs during a certain time period after the scheduled time of arrival (STA). Also, they introduce a 15% capacity buffer to deal with employee illness and higher than expected workloads. A flight arrival delay could cause such an enlarged workload because the maintenance job has to be completed within a reduced time

window. The evaluation of the workforce rosters occurs after simulating flight delays, based on actual arrival data over a complete season. The results retrieved from the simulation model are evaluated with DEA to obtain a number of dominant rosters from which the company can choose.

The remainder of the paper is organized as follows. Section 2 offers a review of the recent literature on decision-support models for stochastic workforce problems. In Section 3 we present our methodology. Section 4 addresses a case study. We introduce the problem setting and give a description of the simulation model and its results. Next, the DEA-model and its results are discussed. Finally, some conclusions and further research perspectives are offered in Section 5.

2 Literature review

In a recent review paper on personnel scheduling literature (Van den Bergh, Beliën, Demeulemeester, De Boeck, & De Bruecker, Unpublished results), we showed that most papers study deterministic problems. Deng and Lin (2011), for instance, propose an ant colony optimization algorithm in order to solve an airline crew scheduling problem. Their goal in choosing this algorithm is to minimize the total crew costs which consist of rest expenses, required pay for each duty and under-utilized time between duties. However, factors such as capacity, amount of workload and/or timing of the workload are subject to variability in real-life situations. Many researchers utilize scenario analysis to test alternative combinations of the exogenous variables in order to create robust schedules. In so doing, the decision makers are provided with the effects of the variability of the decision-support criterion (Borgonovo & Peccati, 2011). Corominas, Lusa and Pastor (2004; 2007) formulate a mixed integer linear programming model to allocate the workforce, considering annualized hours. They use a number of scenarios to test the influence of the demand curves and working-week-type parameters. Brunner and Bard (2012) investigate the problem of determining the size and structure of a workforce that faces time-dependent demand and a variety of labor restrictions and work rules. They address several types of scheduling flexibility in different problem instances, including different shift starting times and lengths, the need for lunch breaks and the option of varying the number of days off.

A second and better way to create robust schedules that can cope with the uncertainty occurring in real-life personnel scheduling problems is to incorporate this uncertainty into the decision-making model. This method is common in call center applications, where the workload depends on the distributions of call arrivals and durations (e.g., Atlason, Epelman, and Henderson (2004, 2008); Avramidis, Chan, Gendreau, L'Ecuyer, and Pisacane (2010); Avramidis, Chan, and L'Ecuyer (2009)). This method can also be used in all types of service organizations. Bard, Morton and Wang (2007), for instance, address a staff planning and scheduling problem that arises at United States Postal Service (USPS) mail processing & distribution centers (P&DCs) and develop a two-stage stochastic

integer program. In the computational phase of the work, three scenarios are considered: high, medium and low demand. The resulting stochastic optimization problem is a large-scale integer program that embodies the full set of contractual agreements and labor rules governing the design of the workforce at a P&DC. The paper of Mason, Ryan and Panton (1998) details a simulation and optimization-based system for the personnel scheduling (rostering) of customs staff at Auckland International Airport, New Zealand. The approach starts with a simulation system embedded in a heuristic search to determine minimum staffing levels. Next, these requirements are used in an integer programming model in order to optimally allocate staff to each period of the working day. These shifts are then assigned to daily work schedules. This method ensures that all passenger processing targets are met by creating high-quality rosters and has resulted in significant labor cost savings. Yeh and Lin (2007) shorten patient waiting times in a hospital emergency department (ED) by adjusting nurses' schedules. They use a simulation model to cover the complete flow for the patient through the ED. The results from the simulation are then fed into a genetic algorithm that makes adjustments to the nurses' schedules. Their results indicate that the average waiting time when using the staffing plan can be reduced significantly.

In reviewing the literature, we noticed that authors are eager to explain the structure of the algorithms developed to create feasible personnel schedules, but only little is known about how good those schedules actually are. Frauendorfer and Königspurger (1996) present a framework for the improvement of (production) scheduling decisions in chemical processing environments using multi-criteria analysis. The purpose of the framework is to provide complete scheduling performance transparency and to contribute to better decision making. Ruiz-Torres, Ho, and López (2006) address a supply chain scheduling problem where both internal and external/outsourced parallel resources are available and the objectives are to minimize the number of late orders and the total outsourced machine time. Several heuristics are developed that generate sets of Pareto-efficient schedules. The evaluation (comparison of the heuristics) is based on two experiments where the values of relevant parameters (e.g., due date tightness or number of jobs) are set at different levels.

In our paper, we want to consider cost minimization and service maximization simultaneously, together with other criteria. One way of dealing with multiple criteria is through multi-objective optimization. Tsai and Li (2009) develop a two-stage mathematical model for a nurse scheduling system. Their genetic algorithm deals with both hospital management requirements and nursing staffs' shift preferences. A different approach is used by Castillo, Joro and Li (2009). They depart from the standard in workforce scheduling to minimize the cost of labor subject to a target service level. They argue that this single operational measure is neither sufficient for capturing the performance of service organizations, nor for characterizing and quantifying service quality. In their approach, they generate a large number of plausible schedules to be evaluated with multiple criteria. Among the evaluated schedules, the efficient ones are identified and the best one is chosen. For the evaluation of these

schedules, they use a Free Disposable Hull algorithm. This approach enabled the management to transform their needs into efficient schedules and to understand the relationship between the cost of labor and various service-quality criteria. The final stage is rather uncommon in the literature. The previously mentioned paper of Yeh and Lin (2007) ends with the remark that their results “can be provided to the hospital administrators for use in decision-making towards enhancing patient care and staffing”. The administrators have to choose between schedules with two shifts and those with three shifts and are provided with simulation results for both alternatives. In our approach we want to facilitate this step by incorporating an evaluation stage for the different personnel rosters.

Instead of a Free Disposable Hull algorithm we will use DEA to evaluate the different rosters. DEA is a method used to determine the relative efficiencies of a set of organizational units, defined as decision making units (DMUs), considering multiple inputs and outputs (Charnes, Cooper, & Rhodes, 1978). Examples of such units to which DEA has been applied are: banks, police stations, hospitals, tax offices, prisons, defense bases (army, navy, air force), schools and university departments. In our study, we will consider the rosters as decision-making units, each with its own inputs and outputs. For the selection of the inputs and outputs and the specific DEA-model, we checked the paper of Dyson et al. (2001). They provide some pitfalls that have been identified in application papers and suggest protocols which can be used in order to avoid these pitfalls and to guide the application of the methodology.

In a DEA study, one wants to project the inefficient DMUs onto the frontier. To do so, there are three different approaches. The input-oriented one aims at reducing the input amounts by as much as possible, while keeping at least the current output levels. The output-oriented one maximizes output levels while maintaining at most the current level of input consumption. The third option deals with the input excesses and output shortfalls simultaneously in a way that maximizes both (Cooper, Seiford, & Tone, 2006, p. 115). When DEA is applied to personnel scheduling problems, it generally focuses on the productivity and efficiency of the workers. Paradi, Smith, and Schaffnit-Chatterjee (2002) use DEA to examine the productivity, efficiency and effectiveness of the Engineering Design Teams (EDT) at Bell Canada. More recently, Azadeh, Ghaderi, Mirjalili, and Moghaddam (2011) have presented an integration of Analytical Hierarchy Process (AHP) and DEA for the assessment and optimization of personnel productivity in a large private bank.

3 Methodology

In this section we present the methodology we will use to select personnel rosters for a given personnel scheduling problem.

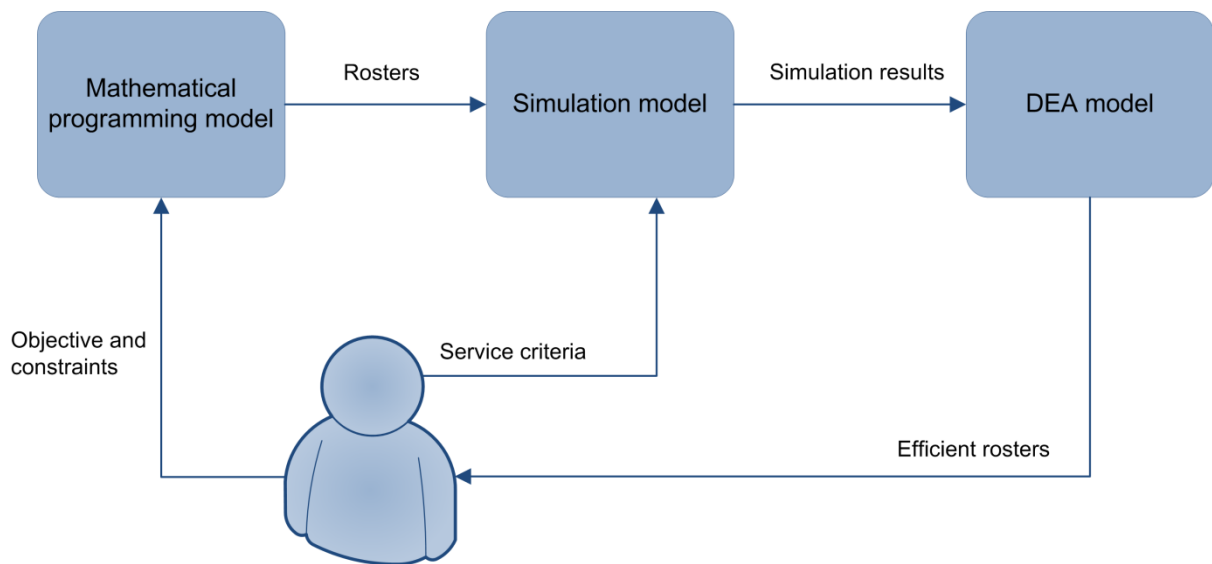


Figure 1: Methodology for selecting personnel rosters

In the first step in our methodology the objective(s), parameters and constraints of the personnel scheduling problem need to be defined. A typical pitfall where this step is concerned is the formulation of an over-constrained model. The developer of the mathematical programming model needs to pursue the management of the company to indicate the most important constraints, such as legal restrictions and minimal coverage constraints. The developer can choose to incorporate the decision on shift definitions into the model or to use predetermined shifts. The typical objective of the model is to minimize costs while satisfying coverage constraints. While the mathematical programming is running, all feasible solutions or rosters are saved. This strategy differs from the common approach of selecting a single best solution. If necessary, the constraints that we left out of the model can be used later on in order to evaluate all these feasible solutions using a simple heuristic.

In the next step, we mimic the behavior of the events in the personnel scheduling problem with a simulation model. Here the goal is to test how the different rosters perform when, for instance, the workload or capacity copes with uncertainty. In order to retrieve meaningful results, the simulation model requires verification and validation. Therefore, one needs to test it intensively with a range of different parameters. It is also important to identify a number of service criteria in coordination with the management of the company. If the first results indicate that the rosters do not meet any of the targets set in the simulation model, it can be necessary to adjust or add to some of the constraints in the mathematical programming model.

The third step consists of treating the rosters as DMUs in a DEA model. The mathematical programming model supplies the DEA model with information on the costs and the shift sequence of the rosters, while the simulation model evaluates the performance on the basis of the service criteria. In this step, one can define relative or absolute restrictions on the weights for the different input or

output factors in line with the preferences of the management or personnel. The DEA model identifies the efficient rosters which are then presented to the management of the company. They can choose which roster to implement, since they have been provided with all the information they need concerning the different steps.

4 Case study

4.1 Problem description

As mentioned in the introduction, we use the rosters built up by Beliën et al. (2012). The objective of the MILP-model is to minimize the labor costs while satisfying a number of constraints, such as coverage constraints. These coverage constraints ensure that for each time period the workload does not exceed the available capacity.

Unlike many other workforce scheduling models, the workload is not given per time period, but per client (flight). Therefore, the number of workers that need to be present per time period depends heavily on the flight schedule. Every flight has a time window, bounded by its STA and STD, during which the maintenance should take place. For each client flight, Sabena Technics has an estimate of the workload, expressed in man-hours. The workload input for their optimization problem is a cyclic list of 310 flights with STA, STD and the estimated workload.

The solution algorithm of Beliën et al. (2012) follows an enumerative approach and solves a MILP for each combination of team sizes for the different cycles. In order to avoid spending MILP time on combinations which cannot result in a better solution, each MILP is provided with a current upper bound (i.e., current best solution). Consequently, whenever the MILP lower bound (which continuously increases during the MILP search) becomes larger than or equal to this upper bound (UB) cut-off value, the MILP calculation stops and the algorithm moves on to the next combination. To allow for several near-optimal solutions, the UB cut-off value can be increased by a certain percentage, e.g., 5%. Every time a MILP results in a feasible solution, it will be saved. At the end of the solution algorithm, multiple feasible solutions might be found, differing in shift sequence, team size, cost, etc.

One constraint in the MILP-approach of Beliën et al. (2012) is the implementation of a time buffer after the flight arrival. This buffer avoids problems in the event of a flight delay. Another constraint adds a time buffer before the flight departure. Since passengers do not like to see aircraft maintenance while they are waiting at the gate, the latest maintenance time is set to 15 or 30 minutes before departure, depending on the available time window. During both time buffers, no maintenance can be scheduled. Both constraint types serve the same goal. They make the work rosters more robust, meaning that small differences in, e.g., workload would not result in huge capacity problems.

However, about one out of seven flights arrives with a delay of more than 15 minutes. Depending on the available maintenance time window and the duration of the time buffers, those delays can lead to coverage problems: the crew member who was assigned to maintain the aircraft could have ended his/her shift, the time window has become too narrow to carry out the maintenance with the pre-assigned number of crew members, etc. On request of the management, the model contains a 15% capacity buffer, added to the coverage constraints. The reason behind this measure is that the company expected problems with the labor unions, because were the new rosters to be introduced, the workload would increase dramatically. This measure would also suffice to deal with situations where the maintenance workload is higher than expected, or capacity is lower than scheduled due to, e.g., employee illness. A flight arrival delay can be seen as a possible cause for the increase in workload. In the event of this happening, the available maintenance time window narrows, which leads to an increase in work pressure.

In the MILP-model, the flights are scheduled in the best possible way. The main principle is to level the workload over the entire time window. When two flights are to be maintained with $STA_1 < STA_2$ and $STD_1 > STD_2$, flight 2 can cause preemption of the maintenance of flight 1 if limited capacity is available, because it arrives later and it needs to depart earlier. Also, the number of workers assigned to a flight can decrease and again increase (and vice versa) during its time window, creating a fluctuating pattern. In real life this kind of behavior would not occur. While scheduling the maintenance, one tries to avoid preemption as well as the assignment of a variable number of workers to a flight. A third drawback of the MILP-approach is the assignment of fractional workers. A job with a workload of one and a half hours can be scheduled during a time window of an hour by assigning 1.5 workers to that job. In practice, one would schedule one worker for the first and an extra worker for the second half hour, or vice versa, or three workers for a half hour, etc.

A simulation procedure was programmed in C++ using Microsoft Visual Studio 2010. A discrete-event simulation method permits us to model real-life behavior, such as arrival delays and the assignment of (discrete) workers to a flight. Real-time flight arrival data for the winter season in 2008 are used in order to identify the appropriate delay distributions. These distributions are applied to the list of flights that was also the input for the MILP-approach (Beliën, et al., 2012). In the simulation, decisions are made every 15 minutes and the planning horizon is divided into quarters of an hour (i.e., each period equals 15 minutes). The available capacity is cyclic for each week. In our case, we simulate a period of 28 weeks, which is the length of an entire flight season with two additional weeks as a warm-up period during which no statistics are collected. Out of the results of the MILP, 50 rosters are selected: 40 rosters with three cycles (i.e., in the third (weekend) cycle the team is only working during special weekend shifts on Friday, Saturday and Sunday and has the rest of the week off) and 10 with only two (regular) cycles. The rosters are selected on the basis of total cost, which is the evaluation criterion in the MILP-approach. The rosters with three cycles can be divided into two

groups, depending on the tightness of the coverage constraints (i.e., the size of the capacity buffer). Out of both groups the 20 cheapest rosters are selected.

We only evaluate the rosters with respect to the effect of uncertainty in the timing of the workload (flight arrival). Other sources of uncertainty, such as available capacity (e.g., illness) or workload (e.g., increased workload for a flight due to a certain failure) are not considered. Within the simulation, the same random numbers are generated in order to evaluate the different scenarios. These differ in terms of the maximum number of workers one can assign to a flight during one period and in the queuing discipline which determines the priority of the flights queuing for line maintenance. We assume that a worker becomes available to work on a new job immediately after finishing a previous one and that every employee is qualified to work on any job (i.e. identically skilled workers). This assumption originates from the MILP-model of Beliën et al. (2012). In real life, however, workers have different skills. The management of the service company considered that using identical workers in the model would not harm feasibility, when team sizes are large enough. For those team sizes, it would be possible to assign workers with different skills to the different teams.

4.2 Simulation results

One of the simulation outputs is the success rate, i.e., the percentage of flights that are maintained before their STD. Table 1 gives the average, minimum and maximum success rates over the 50 rosters with a given maximum number of workers that can be allocated to a certain flight (further referred to as an allocation rule), and different queuing disciplines. We report the average results over all rosters, because these values give us an idea of the performance of the MILP-approach of Beliën et al. (2012) in a stochastic setting. Recall that the solution method of Beliën et al. (2012) does not provide a single “best roster”, but generates many “good acceptable rosters”. Since, in practice, one tries to schedule the maintenance by using a maximum of five workers, we opted for two, three, four and five workers as a maximum number. Four different queuing disciplines are used to order the flights: Earliest STD, Critical ratio, Smallest workload and earliest expected time of arrival (Earliest ETA). For Earliest ETA, we use the expected arrival time, which is the original (deterministic) STA adjusted with a (stochastic) delay. Earliest STD (ETA) establishes the order of flights based on its STD (ETA), starting with the first flight to depart (arrive). The “Critical ratio” uses the division of the workload of a flight by its remaining time interval in order to determine the priority of the flights to be processed. The “Smallest workload” queuing discipline gives priority to the flights that have the smallest workload in man-hours (i.e., not divided by the number of allocated workers). For all the scenarios, using earliest STD as queuing discipline provides the best results for the minima, maxima and averages of the set of rosters. When five workers can be allocated to the same flight, on average 99.44% of all flights leave before their STD.

Worker s	Max											
	Earliest STD			Critical ratio			Smallest workload			Earliest ETA		
	MIN	MAX	AVG	MIN	MAX	AVG	MIN	MAX	AVG	MIN	MAX	AVG
2	0.9009	0.9133	0.9101	0.8886	0.9103	0.9019	0.8967	0.9116	0.9058	0.8842	0.9081	0.8962
3	0.9716	0.9839	0.9810	0.9578	0.9807	0.9720	0.9703	0.9828	0.9772	0.9574	0.9809	0.9691
4	0.9802	0.9925	0.9897	0.9660	0.9891	0.9804	0.9790	0.9913	0.9859	0.9666	0.9896	0.9781
5	0.9849	0.9972	0.9944	0.9705	0.9938	0.9850	0.9832	0.9957	0.9904	0.9714	0.9944	0.9828
AVG	0.9594	0.9717	0.9688	0.9457	0.9685	0.9598	0.9573	0.9704	0.9648	0.9449	0.9683	0.9565

Table 1: Average success rates over the 50 test rosters for different allocation rules and different queuing disciplines

The difference between the minimum and the maximum success rate of the different rosters is only 1.23% for the scenario with a maximum of three workers and earliest STD as queuing discipline. This comes down to a difference of three to four flights per week. Similar results are retrieved with the combinations of two, four and five workers and earliest STD. The difference in costs (wages) between the most expensive and the cheapest roster, however, is more than 26%. Thus, the extra cost of these “expensive” rosters is not automatically translated into a higher success rate. For one specific roster (Roster 50), the contradiction holds that it is by far the most expensive one while having the worst success rate of all the rosters.

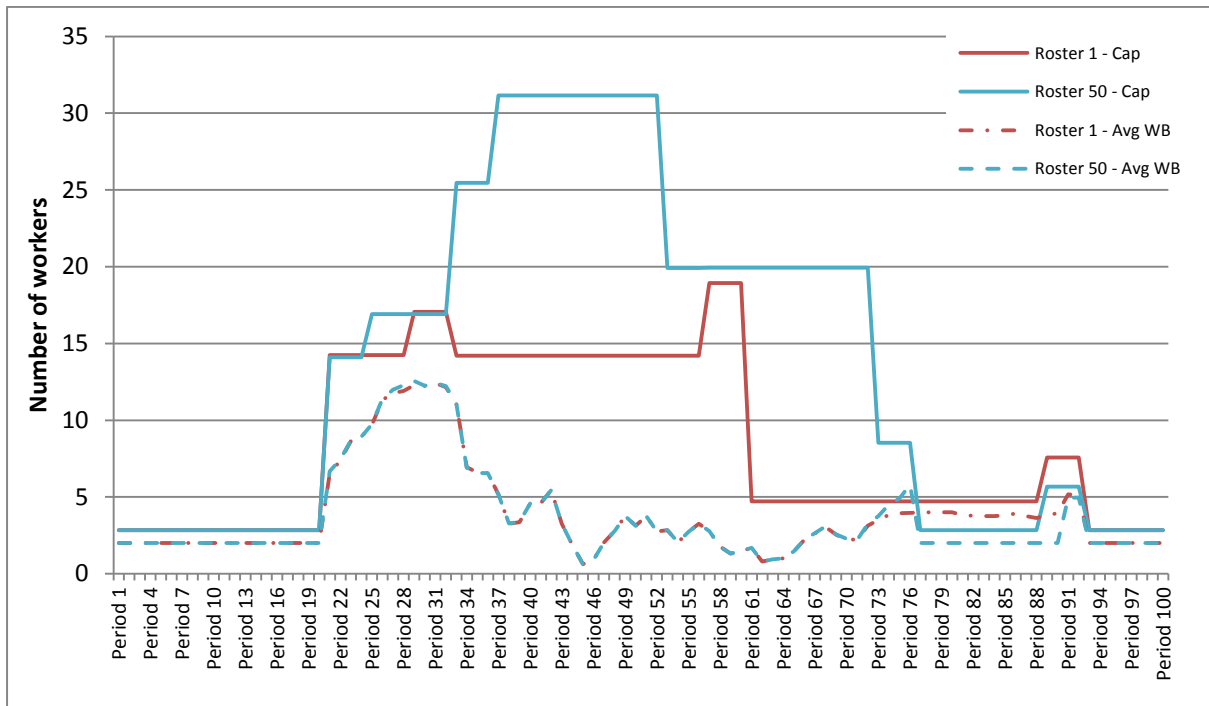


Figure 2: Comparison between the best (Roster 1) and the worst roster (Roster 50) in terms of capacity and average number of workers busy

In Figure 2 the capacity (Cap) and the average number of workers busy (Avg WB) for the first 100 periods (i.e., similar figures could be made for other periods) are shown. Roster 50 clearly has more

capacity during most of the periods, which makes it more expensive. The pattern of the average number of workers busy is similar for the two rosters, as they use the same queuing discipline. The reason why Roster 50 performs worse than Roster 1 is the drop in capacity during specific periods. Due to the lower capacity in these periods (i.e., periods 77 to 93), fewer workers can be allocated to the flights, which results in fewer workers busy. The decrease in performance is caused by a number of flights having their maintenance time window during these periods and now having to wait until other flights are finished or until a new shift starts.

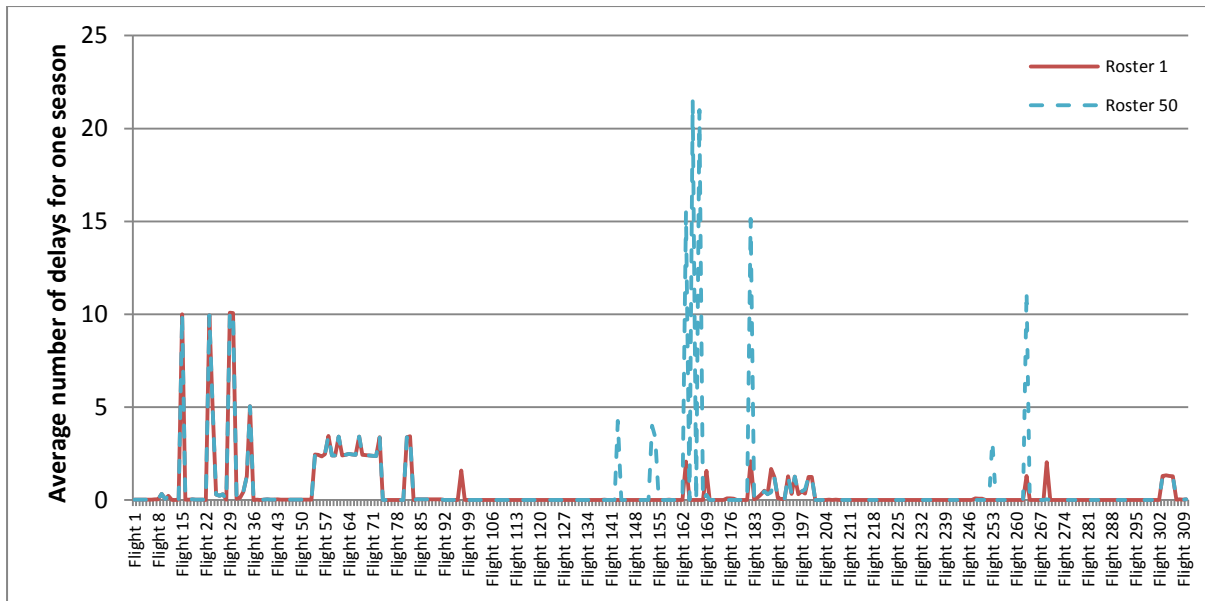


Figure 3: Comparison of the average number of delays between Roster 1 and Roster 50

Figure 3 clearly indicates the difference in the average number of delays between Roster 1 and Roster 50. While there are peaks of up to 10 delays (out of 26 departures) for both rosters, Roster 50 has a number of flights accounting for more than 15 or even more than 20 delays.

So far, we can conclude from Table 1 that the MILP-approach of Beliën et al. (2012) is successful in covering demand under conditions of uncertainty. However, there is still room for improvement. The existence of rather big gaps between the number of workers present and workers busy offers some opportunities to improve the matching between demand and coverage, and hence to cut labor costs.

Besides the results represented in Table 1, we can also calculate the highest attainable success rate a roster can achieve given the maximum number of workers. Since a flight's STD remains fixed in our simulation study, the time window during which maintenance is able to take place can be calculated once the arrival time is known. If the workload of the flight divided by the maximum number of workers is greater than the resulting time window that is left, it is known for certain that the flight cannot depart on time, regardless of the queuing discipline or allocation rule that is used.

Max 2 workers	Max 3 workers	Max 4 workers	Max 5 workers
0.0866	0.0160	0.0074	0.0027

Table 2: Percentage of flights that cannot depart on time for a given maximum number of workers

Table 2 gives an overview of these highest attainable success rates. When the company wants to allocate a maximum of three workers to a flight, the highest achievable success rate is 98.4%. From Table 1 we know that for a maximum of three workers and earliest STD as queuing discipline, the average success rate over the 50 rosters is 98.1%, which represents a difference of only 0.3% or only one flight per week. This strengthens our findings that the MILP-approach of Beliën et al. (2012) is able to cope with the stochastic arrival pattern of the flights. For the remainder of the flights (i.e., those that cannot be handled on time because of the limit on the number of workers per flight), arrangements need to be made between the maintenance company and the airlines. Major delays are generally known beforehand, since the crew can send a telex to the airport of destination. Hence the flights are rescheduled, resulting in a new STD and thereby a new time window for the maintenance company.

4.3 DEA

After defining the rosters by Beliën et al (2012), the next step consists of choosing a roster to implement. So far, the management has had to base its decision on information about the cost of a roster and its shift sequence. Thanks to the simulation experiment, we now have information on the number of flights (and workers) busy during each period, the resulting delays, the number of times a flight has been maintained without preemption, etc. These results aid in searching for the better rosters in the set.

We will now evaluate the rosters by applying DEA. Table 3 lists the inputs and outputs (the original values can be found in Appendix A). The input ‘Cost’ represents the total labor cost of the roster, assuming a fixed fee per hour and an extra premium for the hours worked in early, evening and night shifts and weekends and a fixed premium for every night shift in the roster. For the input ‘Score’, we have developed a tabu search algorithm that takes into account employee preferences when sequencing the different shifts. The MILP-model of Beliën et al. (2012) allocates shifts to every day of the week in a cycle, but does not provide details on the distribution of these shifts over the weeks in the cycle. When, for instance, a cycle has a length of three weeks and two night shifts are assigned to Monday, we have the option between weeks 1 and 2, weeks 1 and 3, and weeks 2 and 3 for assigning the two night shifts. In creating these shift sequences, we still have to satisfy the legal constraints which prohibit an evening shift being followed by a morning or day shift and a night shift being followed by any other shift except for a night shift. The tabu search tries to minimize a penalty cost by swapping two shifts on a certain weekday between the weeks in the cycle. The penalty cost is calculated as follows:

- Invalid shift succession: 1000 points. Neglecting the legal constraints needs to be avoided.
- Incomplete weekends: 100 points. Employees like to work the entire weekend or have the complete weekend off.
- Isolated shift: 50 points. Not only are isolated working shifts unpleasant for workers, but so also are isolated days off.
- Different shift types on Saturday and Sunday: 20 points. On successive weekend working days, the same shift should be assigned.
- Different shift types on weekdays: 10 points. As with the previous penalty, the same shift type should be assigned on successive working days within a week. This creates regular working patterns which are favored by the personnel.
- Block lengths: preferably between [5,8] – if not, a penalty is assigned that proportionally increases, starting from 5 points. If possible, we want to create blocks of minimum 5 working days. Shorter block lengths also cause shorter blocks of days off, which are undesirable.

Inputs	Outputs
Cost	Success rate
Score	Average number of completions without preemption
	Average tardiness

Table 3: Inputs and outputs of the DEA-model

The first output is the success rate of a roster, which represents the average number of flights that were maintained during their time window over all flights. The second output represents the average number of times a flight was maintained without preemption. The third output is the average tardiness. When a flight's maintenance has been finished during its time window, it does not matter whether this was at the beginning or at the end of that interval. Therefore, we only take into account time delays of flights that departed after their STD (i.e., average tardiness) rather than computing the difference between every flight's STD and the end of its maintenance (i.e. average deviation from STD or average lateness).

In our study, we will consider the input-oriented direction in order to define efficient DMUs, since the outputs in our model are provided by the simulation model and our goal is to evaluate the rosters generated by the MILP-model of Beliën et al. (2012). The rosters result in the input variables in the DEA-model. We are interested in identifying the efficient rosters rather than determining how much the cost or score of inefficient rosters should decrease to make them efficient. Moreover, we assume constant returns to scale (i.e., we consider the basic *CCR-I model*), since all rosters satisfy the same MILP-constraints.

Many of the DEA-models assume isotonic inputs and outputs, i.e., increasing the inputs reduces efficiency, whereas efficiency would improve by increasing the outputs. On the basis of this assumption, our inputs are suitable: minimizing both 'Cost' and 'Score' would increase efficiency. The output factors yield some problems, though. The average number of completions without preemption and the success rate should be maximized, whereas the average tardiness should be minimized. These anti-isotonic data are also called *undesirable inputs and outputs*. Dyson et al. (2001) suggest three options to handle anti-isotonic factors: inversion, moving the variable from the output to the input side of the model or subtracting the value of the undesirable factor from a large number.

The transformation of these anti-isotonic variables is one of the pitfalls described by Dyson et al. (2001). The alternative ways of dealing with the problem can lead to different outcomes in terms of the units identified as efficient and the targets set for inefficient units. We tested the influence of two different settings. In the first one, we reverse all the values while in the second one the new values are calculated by subtraction from a large number. Given the allocation rule of a maximum of four workers and earliest STD as queuing discipline, both DEA-models yield the same results: rosters 1, 2, 3, 4, 16 and 35 are considered efficient (see Appendix B).

In the original DEA formulations, the DMUs can freely choose the weights or values to be assigned to each input and output in a way that maximizes their efficiency, subject to this system of weights being feasible for all other DMUs. The advantage of this full flexibility is that it makes clearer the identification of inefficiency, since a DMU is free to choose its own value system. However, an efficient DMU may become so by assigning a zero weight to the inputs and/or outputs for which its performance is worst (Cooper, Seiford, & Zhu, 2004, p. 100). As can be seen in the table in Appendix B, almost all the rosters or DMUs assign a zero weight to at least one of the outputs, while Roster 1 is the only DMU which allocates a zero weight to one of the inputs, i.e., the score of the roster. By doing so, the roster becomes efficient, even with the worst score out of all rosters for the shift sequence (i.e., the score of 1.344 points reveals the existence of an invalid shift succession). We want to avoid neglecting some of the inputs and outputs by adding weight restrictions. There are many ways of adding these weight restrictions: they can be applied directly to the DEA weights or to the product of these weights with the relevant input or output level (i.e., virtual weights). Within these two variations, one can also choose to use absolute or relative weight restrictions (Cooper, et al., 2004, pp. 108-115). Weight restrictions can introduce numerous pitfalls (Dyson, et al., 2001). Podinovski (2004) shows that in the presence of absolute weight bounds, the CCR model and its linear forms may identify the maximum relative efficiency of a DMU incorrectly and he also identifies certain types of restrictions that do not cause the observed error.

We propose the following (absolute) weight restrictions, creating an *assurance region global model* (Cooper, et al., 2006, p. 185) (*AR-Global*):

$$0.75 \leq \frac{\text{Virtual weight of Cost}}{\text{Total virtual weight}} \leq 0.95 \quad (1)$$

$$0.05 \leq \frac{\text{Virtual weight of Score}}{\text{Total virtual weight}} \leq 0.25 \quad (2)$$

$$0.1 \leq \frac{\text{Virtual weight of Average number of completions without preemption}}{\text{Total virtual weight}} \leq 0.2 \quad (3)$$

$$0.25 \leq \frac{\text{Virtual weight of Average tardiness}}{\text{Total virtual weight}} \leq 0.5 \quad (4)$$

$$0.25 \leq \frac{\text{Virtual weight of Success rate}}{\text{Total virtual weight}} \leq 0.5 \quad (5)$$

Restrictions (1) and (2) emphasize the relative importance of the input ‘Cost’ as compared to the input ‘Score’. For the maintenance company, the main goal was to cut labor costs while satisfying demand. The rosters needed to be feasible and if possible, meet personnel requests. Therefore, we stress the importance of the input ‘Cost’, while the low virtual weight for the input ‘Score’ makes it possible to differentiate between those rosters with valid and those with invalid shift sequences and to a lesser extent, those with preferred shift sequences. Restrictions (3), (4) and (5) represent the weight restrictions on the output variables. It is clear that the outputs ‘Average tardiness’ and ‘Success rate’ are more important than the output ‘Average number of completions without preemption’.

After adding these weight restrictions, only three rosters appear to be efficient: Rosters 3, 16 and 35 (see Appendix C). Of the efficient rosters, Roster 35 is the only roster that does not set the weight of the input ‘Cost’ at the upper bound of 95%, but emphasizes its performance on the input ‘Score’ instead. This is logical, since Roster 35 has the minimum value for this input variable. The weights assigned to the output variables are quite similar for the different efficient rosters.

5 Conclusion and future work

Throughout this paper, we discussed a new methodology for selecting personnel rosters. Instead of searching for an optimal roster with a multi-objective mathematical programming approach, we presented a framework of mathematical programming, discrete event simulation and DEA analysis. With the DEA analysis, it is easy to compare the different rosters considering different weight restrictions. In a ‘traditional approach’, this would be rather time-consuming, since it would be necessary to run the mathematical programming model for each setting.

Considering the case study, the simulation model provides insight into the performance of the rosters, resulting from the deterministic MILP-model. In general, the MILP approach performs well. With an easily applicable queuing discipline such as earliest STD, on average only four flights per week out of the cyclic list of 310 flights show departure delays. However, there is still room for improvement. First, we noted that the extra capacity, added to please the unions, is not always assigned to the right periods, which results in a large number of idle workers after the daily morning working peaks. Second, arrival delays and the upper bound on the number of workers that are to be assigned to the same flight (which were not incorporated into the MILP-approach) results in departure delays that cannot be avoided without an adjustment of flights' STD. More than 80% of the departure delays are caused by the limit on the number of workers allocated to each flight. In practice, the company could reconsider the flights' STD, since they are informed of arrival delays by the airlines. We could try to cope with this by adding extra constraints to the MILP-model, considering the restriction on the number of workers as well as preemption, which is still present because of capacity dropdowns during a flight's maintenance time window. The DEA-model allows the maintenance company to select a roster which dominates others based on objective inputs and outputs from the MILP, the tabu search and the simulation model, and by defining specific weights to these input and output parameters. This prevents the selection of a roster on the basis of intuition alone with only the costs and a possible shift sequence as decision parameters.

In future, we can also incorporate other sources of uncertainty into the model, such as uncertainty in supply (e.g., employee illness) and uncertainty in demand (e.g., unexpected aircraft problems leading to higher workloads). A last topic that could be of interest for future research is the incorporation of skills.

Acknowledgements

This research was sponsored by the Special Research Fund (BOF) HUB-KUB.

Appendix A

	Cost	Score	Avg tardiness	Success rate	Avg # of completions without preemption
Roster 1	31671.62	1344	1.7961	0.9911	0.8883
Roster 2	31939.94	211	1.8504	0.9917	0.8930
Roster 3	32051.80	123	1.8470	0.9913	0.8844
Roster 4	32271.90	96	2.0123	0.9892	0.8938
Roster 5	32418.94	1277	1.8643	0.9888	0.8820
Roster 6	32608.72	304	2.0293	0.9890	0.9003
Roster 7	32809.85	552	1.9265	0.9880	0.8931

Roster 8	33045.05	183	1.8406	0.9911	0.9042
Roster 9	33223.98	304	1.8690	0.9914	0.9136
Roster 10	33328.50	473	1.8878	0.9850	0.8895
Roster 11	33435.36	219	1.9670	0.9891	0.8955
Roster 12	33934.50	111	1.8972	0.9919	0.9238
Roster 13	33982.30	344	1.8567	0.9889	0.8839
Roster 14	33991.60	220	1.8417	0.9890	0.8811
Roster 15	34004.81	248	1.9543	0.9861	0.8979
Roster 16	34221.20	68	1.7221	0.9911	0.9271
Roster 17	34323.30	233	1.7637	0.9911	0.9157
Roster 18	34529.51	171	1.8200	0.9902	0.9142
Roster 19	34682.00	228	1.9319	0.9915	0.9255
Roster 20	34827.95	157	1.9495	0.9881	0.8845
Roster 21	32003.25	1154	1.8163	0.9911	0.8887
Roster 22	32034.15	404	1.8904	0.9879	0.8917
Roster 23	32577.45	129	2.0293	0.9890	0.8999
Roster 24	33079.16	353	1.8212	0.9905	0.9110
Roster 25	33416.55	83	2.0219	0.9892	0.8991
Roster 26	33519.61	352	1.8246	0.9912	0.9132
Roster 27	33529.79	1070	1.9895	0.9890	0.8925
Roster 28	33547.30	520	1.9635	0.9893	0.9074
Roster 29	33612.90	501	1.9158	0.9893	0.9064
Roster 30	33905.20	119	1.9254	0.9919	0.9233
Roster 31	34134.71	107	1.8281	0.9919	0.9295
Roster 32	34189.50	144	1.9391	0.9923	0.9062
Roster 33	34267.22	180	1.8220	0.9911	0.9082
Roster 34	34298.85	136	1.9621	0.9923	0.9084
Roster 35	34309.92	45	1.8271	0.9881	0.8962
Roster 36	34578.58	87	1.7640	0.9911	0.9311
Roster 37	35144.94	1123	1.9848	0.9803	0.8905
Roster 38	35212.41	250	1.9079	0.9892	0.9134
Roster 39	35607.09	110	1.9585	0.9925	0.9033
Roster 40	35614.42	107	1.8875	0.9919	0.9241
Roster 41	37165.01	79	1.9354	0.9891	0.8953
Roster 42	37226.21	69	1.8794	0.9922	0.8976
Roster 43	37617.69	155	1.9786	0.9862	0.8985
Roster 44	37724.65	213	2.0193	0.9925	0.9075
Roster 45	37730.13	156	1.9031	0.9916	0.9227
Roster 46	38081.81	92	1.8893	0.9890	0.9059
Roster 47	38467.74	71	1.9545	0.9893	0.8897
Roster 48	38747.62	249	1.9028	0.9889	0.9181
Roster 49	39427.35	312	1.9109	0.9912	0.9426
Roster 50	40150.61	79	2.0617	0.9802	0.8925

Table A.1: Original values of input and output variables for the DEA-model given a maximum number of four workers per flight and earliest STD as queuing discipline

Appendix B

DMU	Score	VX(1) Cost	VX(2) Score	UY(1) Avg # of completions without preemption	UY(2) Avg tardiness	UY(3) Success rate
Roster 1	1	1	0	1	0	0
Roster 2	1	0.999432	5.68E-04	1	0	0
Roster 3	1	0.996324	3.68E-03	6.93E-02	0.860662	7.00E-02
Roster 4	1	0.992232	7.77E-03	1	0	0
Roster 5	0.975338	0.991439	8.56E-03	0	0	0.975338
Roster 6	0.987252	0.999199	8.01E-04	0.987252	0	0
Roster 7	0.972668	0.998555	1.44E-03	0.972668	0	0
Roster 8	0.982682	0.975649	2.44E-02	0.903893	7.88E-02	0
Roster 9	0.98325	0.999214	7.86E-04	0.98325	0	0
Roster 10	0.953867	0.998781	1.22E-03	0.953867	0	0
Roster 11	0.957998	0.983055	1.69E-02	0.957998	0	0
Roster 12	0.983525	0.985472	1.45E-02	0.908341	7.52E-02	0
Roster 13	0.936465	0.997786	2.21E-03	0	0	0.936465
Roster 14	0.942759	0.99365	6.35E-03	0	0.854612	8.81E-02
Roster 15	0.944363	0.999373	6.27E-04	0.944363	0	0
Roster 16	1	0.954985	4.50E-02	1	0	0
Roster 17	0.971624	0.994018	5.98E-03	0.195384	0.77624	0
Roster 18	0.953314	0.978168	2.18E-02	0.87686	7.65E-02	0
Roster 19	0.954452	0.982993	1.70E-02	0.954452	0	0
Roster 20	0.916862	0.986116	1.39E-02	0.457348	0	0.459514
Roster 21	0.991041	0.992157	7.84E-03	0	0	0.991041
Roster 22	0.995006	0.998917	1.08E-03	0.995006	0	0
Roster 23	0.994763	0.989687	1.03E-02	0.994763	0	0
Roster 24	0.984618	0.999083	9.17E-04	0.984618	0	0
Roster 25	0.986206	0.911549	8.85E-02	0.611461	0	0.374744
Roster 26	0.974079	0.999098	9.02E-04	0.974079	0	0
Roster 27	0.949954	0.997263	2.74E-03	0.949954	0	0
Roster 28	0.966696	0.998669	1.33E-03	0.966696	0	0
Roster 29	0.963759	0.99872	1.28E-03	0.963759	0	0
Roster 30	0.981847	0.990848	9.15E-03	0.981847	0	0
Roster 31	0.988121	0.973483	0.026517	0.874699	0.113422	0
Roster 32	0.953838	0.989037	1.10E-02	0.953838	0	0
Roster 33	0.953461	0.976874	2.31E-02	0.876609	7.69E-02	0
Roster 34	0.953735	0.989673	1.03E-02	0.953735	0	0
Roster 35	1	0.927964	7.20E-02	1	0	0
Roster 36	0.986082	0.978604	2.14E-02	0.869436	0.116646	0
Roster 37	0.904255	0.99726	2.74E-03	0.904255	0	0
Roster 38	0.927691	0.99939	6.10E-04	0.927691	0	0
Roster 39	0.91674	0.973856	2.61E-02	0.815207	0.101533	0
Roster 40	0.939773	0.974556	2.54E-02	0.83437	0.105404	0

Roster 41	0.895867	0.920384	7.96E-02	0	0	0.895867
Roster 42	0.906496	0.929862	7.01E-02	0	0	0.906496
Roster 43	0.859822	0.989273	1.07E-02	0.859822	0	0
Roster 44	0.862576	0.985359	1.46E-02	0.862576	0	0
Roster 45	0.880259	0.989236	1.08E-02	0.880259	0	0
Roster 46	0.871505	0.91376	8.62E-02	0.541928	0	0.329577
Roster 47	0.874902	0.930137	6.99E-02	0	0	0.874902
Roster 48	0.847799	0.983369	1.66E-02	0.847799	0	0
Roster 49	0.854938	0.99932	6.80E-04	0.854938	0	0
Roster 50	0.829502	0.92862	7.14E-02	0.514649	0	0.314853

Table B.1: Outcome of the CCR-I-model (normalized data), representing the scores of the rosters and their weights

Appendix C

DMU	Score	VX(1) Cost	VX(2) Score	UY(1) Avg # completions without preemption	UY(2) Avg tardiness	UY(3) Success rate
Roster 1	0.975932	0.95	5.00E-02	0.174612	0.383316	0.418003
Roster 2	0.983972	0.95	0.05	0.166891	0.199353	0.617728
Roster 3	1	0.95	0.05	0.178626	0.392701	0.428673
Roster 4	0.99679	0.928738	7.13E-02	0.141889	0.189626	0.665275
Roster 5	0.939611	0.95	0.05	8.41E-02	0.228365	0.627187
Roster 6	0.937667	0.95	5.00E-02	0.159309	0.180036	0.598322
Roster 7	0.930823	0.95	5.00E-02	0.157655	0.184174	0.588994
Roster 8	0.960056	0.95	5.00E-02	0.172959	0.377479	0.409618
Roster 9	0.939674	0.95	0.05	0.161487	0.18944	0.588748
Roster 10	0.919692	0.95	0.05	0.155505	0.18412	0.580067
Roster 11	0.927634	0.95	0.05	0.157253	0.181342	0.589038
Roster 12	0.959074	0.924705	7.53E-02	0.165841	0.191763	0.60147
Roster 13	0.911422	0.95	5.00E-02	0.153627	0.184207	0.573588
Roster 14	0.923234	0.95	0.05	0.164655	0.362161	0.396418
Roster 15	0.909161	0.95	0.05	0.154411	0.178428	0.576322
Roster 16	1	0.95	5.00E-02	0.181921	0.393974	0.424105
Roster 17	0.935847	0.95	0.05	0.169428	0.368237	0.398182
Roster 18	0.928144	0.95	5.00E-02	0.167919	0.36504	0.395185
Roster 19	0.903773	0.95	5.00E-02	0.156359	0.178828	0.568586
Roster 20	0.901771	0.95	0.05	0.125047	0.174641	0.602083
Roster 21	0.963148	0.95	5.00E-02	0.172356	0.37831	0.412481
Roster 22	0.960426	0.95	5.00E-02	0.162624	0.192138	0.605664
Roster 23	0.965445	0.95	0.05	0.163981	0.185364	0.6161
Roster 24	0.94779	0.95	0.05	0.17124	0.372737	0.403812
Roster 25	0.973993	0.939808	6.02E-02	0.139137	0.184796	0.650059
Roster 26	0.935649	0.95	5.00E-02	0.169212	0.368117	0.39832
Roster 27	0.900137	0.95	0.05	0.127001	0.172339	0.600796

Roster 28	0.911018	0.95	0.05	0.155684	0.178392	0.576942
Roster 29	0.914004	0.95	5.00E-02	0.156213	0.181561	0.57623
Roster 30	0.951728	0.919686	8.03E-02	0.164444	0.188666	0.598618
Roster 31	0.964716	0.92755	7.25E-02	0.167618	0.197155	0.599943
Roster 32	0.92857	0.95	5.00E-02	0.158614	0.18315	0.586806
Roster 33	0.932025	0.95	0.05	0.168196	0.36655	0.397279
Roster 34	0.926452	0.95	0.05	0.158424	0.181502	0.586526
Roster 35	1	0.933862	6.61E-02	0.179517	0.392544	0.427939
Roster 36	0.973176	0.940935	5.91E-02	0.169435	0.203072	0.600669
Roster 37	0.853963	0.95	0.05	0.144238	0.166013	0.543712
Roster 38	0.887158	0.95	0.05	0.152342	0.176716	0.5581
Roster 39	0.909088	0.925947	7.41E-02	0.127353	0.175674	0.606061
Roster 40	0.92044	0.930403	6.96E-02	0.159214	0.184591	0.576635
Roster 41	0.890623	0.947675	5.23E-02	0.122969	0.173217	0.594436
Roster 42	0.903498	0.941668	5.83E-02	0.122338	0.178771	0.602389
Roster 43	0.836988	0.95	0.05	0.142152	0.163083	0.531753
Roster 44	0.825411	0.95	5.00E-02	0.140941	0.159008	0.525463
Roster 45	0.847536	0.95	0.05	0.146432	0.169148	0.531956
Roster 46	0.868554	0.940661	5.93E-02	0.11755	0.171283	0.579721
Roster 47	0.864921	0.941437	0.058563	0.120403	0.16728	0.577238
Roster 48	0.809516	0.95	0.05	0.139446	0.161539	0.508531
Roster 49	0.795022	0.95	0.05	0.139105	0.158445	0.497472
Roster 50	0.812274	0.9375	6.25E-02	0.116224	0.152285	0.543765

Table C.1: Outcome of the AR-Global model (normalized data), representing the scores of the rosters and their weights

References

- Atlason, J., Epelman, M. A., & Henderson, S. G. (2004). Call center staffing with simulation and cutting plane methods. *Annals of Operations Research*, 127, 333-358.
- Atlason, J., Epelman, M. A., & Henderson, S. G. (2008). Optimizing call center staffing using simulation and analytic center cutting-plane methods. *Management Science*, 54, 295-309.
- Avramidis, A. N., Chan, W., Gendreau, M., L'Ecuyer, P., & Pisacane, O. (2010). Optimizing daily agent scheduling in a multiskill call center. *European Journal of Operational Research*, 200, 822-832.
- Avramidis, A. N., Chan, W., & L'Ecuyer, P. (2009). Staffing multi-skill call centers via search methods and a performance approximation. *IIE Transactions*, 41, 483-497.
- Azadeh, A., Ghaderi, S. F., Mirjalili, M., & Moghaddam, M. (2011). Integration of analytic hierarchy process and data envelopment analysis for assessment and optimization of personnel productivity in a large industrial bank. *Expert Systems with Applications*, 38, 5212-5225.
- Bard, J. F., Morton, D. P., & Wang, Y. M. (2007). Workforce planning at USPS mail processing and distribution centers using stochastic optimization. *Annals of Operations Research*, 155, 51-78.

- Beliën, J., Demeulemeester, E., & Cardoen, B. (2012). Improving workforce scheduling of aircraft line maintenance at Sabena Technics. *Interfaces*, 42, 352-364.
- Borgonovo, E., & Peccati, L. (2011). Managerial insights from service industry models: a new scenario decomposition method. *Annals of Operations Research*, 185, 161-179.
- Brunner, J., & Bard, J. (2012). Flexible weekly tour scheduling for postal service workers using a branch and price. *Journal of Scheduling*, In press.
- Castillo, I., Joro, T., & Li, Y. Y. (2009). Workforce scheduling with multiple objectives. *European Journal of Operational Research*, 196, 162-170.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429-444.
- CODA. (2010). Delays to air transport in Europe. In *CODA Digest* (pp. 44): Eurocontrol.
- Cooper, W. W., Seiford, L. M., & Tone, K. (2006). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*: Springer.
- Cooper, W. W., Seiford, L. M., & Zhu, J. (2004). *Handbook on Data Envelopment Analysis*: Springer.
- Corominas, A., Lusa, A., & Pastor, R. (2004). Planning annualised hours with a finite set of weekly working hours and joint holidays. *Annals of Operations Research*, 128, 217-233.
- Corominas, A., Lusa, A., & Pastor, R. (2007). Using a MILP model to establish a framework for an annualised hours agreement. *European Journal of Operational Research*, 177, 1495-1506.
- Deng, G.-F., & Lin, W.-T. (2011). Ant colony optimization-based algorithm for airline crew scheduling problem. *Expert Systems with Applications*, 38, 5787-5793.
- Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132, 245-259.
- Eurocontrol. (2011). Planning for delay: influence of flight scheduling on airline punctuality. In *Eurocontrol trends in air traffic* (Vol. 7, pp. 124).
- Frauendorfer, K., & Königspurger, E. (1996). Concepts for improving scheduling decisions: An application in the chemical industry. *International Journal of Production Economics*, 46-47, 27-38.
- Gupta, P., Bazargan, M., & McGrath, R. N. (2003). Simulation model for aircraft line maintenance planning. In *Annual Reliability and Maintainability Symposium, 2003* (pp. 387-391). Tampa, Florida.
- Mason, A. J., Ryan, D. M., & Panton, D. M. (1998). Integrated Simulation, Heuristic and Optimisation Approaches to Staff Scheduling. *Operations Research*, 46, 161-175.
- Paradi, J. C., Smith, S., & Schaffnit-Chatterjee, C. (2002). Knowledge worker performance analysis using DEA: an application to engineering design teams at Bell Canada. *IEEE Transactions on Engineering Management*, 49, 161-172.
- Podinovski, V. V. (2004). Suitability and redundancy of non-homogeneous weight restrictions for measuring the relative efficiency in DEA. *European Journal of Operational Research*, 154, 380-395.
- Ruiz-Torres, A. J., Ho, J. C., & López, F. J. (2006). Generating Pareto schedules with outsource and internal parallel resources. *International Journal of Production Economics*, 103, 810-825.
- Tsai, C.-C., & Li, S. H. A. (2009). A two-stage modeling with genetic algorithms for the nurse scheduling problem. *Expert Systems with Applications*, 36, 9506-9512.

- Van den Bergh, J., Beliën, J., Demeulemeester, E., De Boeck, L., & De Bruecker, P. (Unpublished results). Personnel scheduling: a literature review.
- Yan, S., Yang, T. H., & Chen, Y. C. (2004). A model and a solution algorithm for airline line maintenance manpower supply planning with multiple aircraft type maintenance certificates. *Journal of the Chinese Institute of Engineers*, 27, 719-729.
- Yeh, J.-Y., & Lin, W.-S. (2007). Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department. *Expert Systems with Applications*, 32, 1073-1083.