Cockpit crew pairing Pareto optimisation in a budget airline

Parames Chutima[a,b*] and Nicha Krisanaphan[a]

[a]Department of Industrial Engineering, Chulalongkorn University, Thailand
[b]The Royal Society of Thailand, Thailand

CHRONICLE	ABSTRACT
	Crew pairing is the primary cost checkpoint in airline crew scheduling. Because the crew cost comes second after the fuel cost, a substantial cost saving can be gained from effective crew pairing. In this paper, the cockpit crew pairing problem (CCPP) of a budget airline was studied. Unlike the conventional CCPP that focuses solely on the cost component, many more objectives deemed to be no less important than cost minimisation were also taken into consideration. The adaptive non-dominated sorting differential algorithm III (ANSDE III) was proposed to optimise the CCPP against many objectives simultaneously. The performance of ANSDE III was compared against the NSGA III, MOEA/D, and MODE algorithms under several Pareto optimal measurements, where ANSDE III outperformed the others in every metric.

1. Introduction

Airline scheduling is one of the most challenging planning processes encountered by the airline industry. Due to their complexities, four main aspects of airline scheduling are often sequentially optimised, these being the flight schedule, fleet assignment, aircraft routing and crew scheduling problems (Gopalakrishnan & Johnson, 2005). The result of the preceding problem becomes the given input for the succeeding problem. To enhance their competitive advantages, airlines must optimise the crew scheduling since it is under their control and the total crew costs (e.g. benefits, salary and top-up payment) are the second biggest operating expenditure for airlines after fuel costs. Substantial annual returns can be earned from even a marginal saving percentage from a well-pondered airline scheduling management (Anbil *et al.*, 1991). Airline crew scheduling consists of the two sub-problems of crew pairing and crew rostering, which are normally solved successively. The former determines the set of flight sequences with a minimum cost to encompass a predefined set of flight legs, as prescribed in the timetable of the airline for a given period while abiding by the diverse and complex legal restrictions enacted by the federation's safety regulations and the company's agreements. Under this context, a pairing means a crew itinerary or a tour of duty to be followed by a crew member. The latter pertains to the assignment of named crew members to the optimised crew pairings obtained from the previous step. The assignment priority often depends on senioritis and crew preferences. This paper focuses on the former problem. Crew members in an airline are classified as cockpit and cabin crews (Graves *et al.*, 1993). Cockpit crews (e.g. captain, first officer, etc.) are licensed to fly only a specific aircraft type, depending on their current licenses and certifications. Moreover, different categories of cockpit crews may be controlled by dissimilar permissible restrictions. In contrast, cabin crews (e.g. steward and stewardess) may provide services on more than one type of aircraft since their licenses are less restrictive. In this paper, we confined our research scope to the cockpit crew, hereafter called the cockpit crew pairing problem (CCPP), since airlines pay substantially higher wages and fringe benefits to the cockpit crew than to the cabin crew. Due to the restrictions in the pilot's license, the CCPP is optimised only for an individual fleet type. Besides the economic considerations, the optimised CCPP would bring about increasing cockpit crew utilisation and reducing wastage time due to other reasons apart from flights (Zeren & Özkol, 2016).

Although a crew pairing problem (CPP) is not new, the following reasons render its continued challenge to academia and practitioners. A tremendous number of pairing combinations are possible for most airline companies. Multifaceted and nonlinear functions have to be used to formulate crew costs with a minimum guaranteed pay. Numerous regulations and practical restrictions (e.g. international safety regulations, work rules, labour laws, and airline-specific rules) must be obeyed for feasible pairings. Moreover, nowadays, more and more people are turning to travel by plane causing the air-travel demand (problem size) to increase exponentially to match the rapid expansions of flight routes and frequencies. These problems are acknowledged by all airlines as making a manual generation of crew pairings an ineffective approach. Operations research techniques have been used to solve the CPP since 1950 (Barnhart & Talluri, 1997). Set covering and set partitioning methods have often been exercised in the study of the CPP. Since the CPP falls in an NP-hard class of combinatorial optimization problems, for real-life problems, these techniques are unable to effectively optimise such an intractable problem in reasonable computation time (Deveci & Demirel, 2018). To help find a good solution, many metaheuristics have been developed recently. Most research has focussed only on the single objective of minimising the crew pairing cost (CPC). However, other objectives that are no less important than the cost-related issue, such as workload balancing, should also be taken into account when forming crew pairings. Hence, the objective of this paper was to propose a new approach to solve the multi-objective (MO)CCPP.

The structure of this paper is formatted as follows. The survey of the CPP literature is presented in Section 2. Section 3 explains the background of the CPP in a budget airline case study in Thailand, and then the methodology to optimise the MOCCPP is proposed in Section 4. The experimental cases and benchmarked algorithms are described in Section 5, while the experimental results and analysis are discussed in Section 6. Finally, Section 7 summarises the conclusion of this study.

2. Literature survey

A crew pairing consists of a set of flight legs within the same fleet, operated by the same crew and organised in a sequential order, where the same crew base is destined as the starting and terminating points. During the last 30 decades, the CPP has continued to be challenging due to the size and complexity of the problem. A recent survey of the CPP that discussed the models and solution has been presented (Gopalakrishnan & Johnson, 2005). Lavoie *et al.* (1988) developed a new method to solve the CPP with two steps. First, the set covering problem represented by a 0-1 integer linear program was used to formulate the CPP with the aim of minimising the total CPC. Then, the continuous relaxation of the problem was solved by the column generation (CG) method. This approach was then applied to real long- and medium-haul networks of Air France, with 4 to 5% average cost savings being found over the best manual computations exercised by experts. Barnhart *et al.* (1995) proposed the deadhead selection heuristic to decrease the extended rest periods of crews leading to an overall cost reduction in long-haul international CPP. The profiles of the arrival and departure times at each station, which were used in successively pricing out potential deadhead flights, were the dual solutions obtained from solving the linear programming relaxation. A substantial reduction in the CPC (around $5 million yearly) was revealed from applying the proposed mechanism. Desaulniers *et al.* (1997) used an integer, nonlinear multi-commodity network flow model with supplemented resource variables to formulate the CPP to minimise the CPCs. Integer solutions of the problem were derived from the branch-and-bound (B&B) algorithm in which the concept of the Dantzig-Wolfe decomposition technique was applied. The proposed method was then tested on the medium-haul networks of Air France and was found to be much more effective than the expert system. Anbil *et al.* (1998) applied the sprint method to solve linear programming relaxations and a branch on follow-on tactic in the B&B approach to significantly shorten the solution time in reaching approximate solutions of the CPP. The data from Southwest Airlines and US Airways were then used to test the proposed approach, where the CPU times were reduced dramatically compared with the dual simplex algorithm. Klabjan *et al.* (2001)] proposed an algorithm for the CPP that was composed of two phases. The linear programming relaxation was solved to obtain quasi optimality in the first phase. Then an integer solution with the greatest improved cost was heuristically attained in the second phase. Overall, the algorithm was claimed to perform better than the current practice. AhmadBeygi *et al.* (2009) applied a network flow concept in formulating a mixed-integer programming model for the CPP. The marker and connection variables were employed to model the non-linear cost function and complicated restrictions about crew-to-flight assignments. Under such an approach, the 143- and 329-flight networks operated by a major US carrier could be solved by commercial solvers. Liu *et al.* (2009) created an opportunity to explore more optimal solutions of the CPP by applying the permutation-based mode in the formulation stage. The problem was then solved by the inequalities-based MO genetic algorithm (GA). Several objectives were formalized through mathematical equations. The effectiveness of the proposed method was tested with real data provided by a Taiwanese airline. Souai and Teghem (2009) proposed an integrated model to solve both airline crew pairing and rostering sub-problems simultaneously. A hybrid GA was applied to optimise the bi-objective problem hierarchically, which was to minimise both the total cost and the deviation between the real and average flying times of any crew member. Three real cases from airlines were used to test the performance of the proposed algorithm. Zeren and Özkol (2012) applied a GA with a perturbation genetic operator (GO) to solve the CPP with multiple crew bases and deadhead flights. The test data were obtained from Turkish airlines. The proposed algorithm was then compared with another four approaches and was found to be effective with a fast convergence time and could reduce deadhead flights. Reisi and Moslehi (2013) addressed the daily CCPP, in which deadheading was not permitted. Based on the shortest path with resource constraint, two approaches were developed to solve the CG sub-problem, namely the SPRCD and SPRCF. The data from two major Iranian airlines were employed to test the

performance of the proposed approaches against an existing method (MKA). The solution times of both the SPRCD and SPRCF approaches were substantially lower than MKA. Aydemir-Karadag *et al.* (2013) developed two algorithms for solving the daily CCPP, i.e. a knowledge-based random algorithm (KBRA) and a hybrid algorithm (HA). These two algorithms were compared with the CG technique using hypothetical problems with two home bases. The average crew schedule costs produced by the HA were the same or close to those from the CG approach, and they were better than the KBRA. However, the CPU times of the HA were substantially shorter than the CG. Zeren and Özkol (2016) proposed a new strategy to solve the CCPP that consisted of new CG, paring enumeration, and pricing network methods. The proposed strategy was evaluated with the data from Turkish airlines. Besides obtaining comparative results with the current practice, it successfully reduced the number of deadheads and international overnights (Zeren & Özkol, 2016). Demirel and Deveci (2017) developed a dynamic-based GA for medium-scaled CPP. The deadhead minimising search and partial solution search approaches were also proposed to effectively improve the quality pairings. The proposed approach could find competitive results compared with the current method. Deveci and Demirel (2018) developed a memetic algorithm (MA) with hill climbing and local optimization mechanisms and two variations of the GA to solve the monthly CPP of a Turkish domestic airline. The proposed MA obtained lower CPCs compared with two GAs. Aggarwal *et al.* (2019) presented a customised GA in which the initialisation mechanism and GOs were improved to increase the exploration of search space. The proposed approach targeted optimising the CPC and reducing deadheads and overnight rests of the medium-scaled US airline networks. The performance gap between the proposed algorithm and the CG optimiser was observed to be around 11%.

From the literature survey, the CPP was shown to be a complex problem due to the non-linear CPC function. Most papers have attempted to optimise only a single objective function. Since the CCP is NP-hard, conventional analytical approaches have been unable to optimise practical sized problems. As a result, various heuristics based on mathematical relaxations and evolutionary algorithms (EAs), such as GA, were proposed to find a good and acceptable solution. However, apart from the CPP, which is the most addressed issue, other objectives also play a significant role in the airline industry resulting in MO optimisation of the CPP being a more sensible model. Moreover, many novel MOEAs have emerged recently, which enable finding the Pareto optimal solutions for the MOCCPP effectively and being something that has more practical value.

3. Problem definition

The case study airline is a budget carrier in Thailand that was registered in the securities exchange of Thailand nearly a decade ago. Although the airline is a public company limited, in which many public and private investors are joint ventures, its major shareholder is the biggest national airline enterprise. The mission of the airline is to be the first choice among budget national airlines by paying the greatest endeavours to understand the passengers' demands and needs as well as delivering beyond expectation services to them. The airline offers nearly 30 flight routes, most of which are operated domestically. Nearly 700 roundtrip flights are scheduled weekly using Boeing 737-800, ATR72-500, and Bombardier Q400 aircraft. The main operations base is located at Don Mueang international airport. For a better understanding in the context of airlines, some terminologies often used in addressing the CCPP need to be articulated before gearing towards the remaining discussion (see Fig. 1 for pictorial expressions of these terminologies).

Flight leg (segment): A one-way nonstop flight of the aircraft consisted of take-off at the departure airport, flying to the destination, and landing at the destination airport.

Block hour: The interval from the instance that the doors of the aircraft close at the departure airport until the instance that the doors of the aircraft open at the arrival gate of the destination airport after its landing.

Duty period: A sequence of flight legs with short rest periods (sit times) interposed between them. In addition, brief (preparation) and debrief (end) periods are placed before the first flight leg and after the last flight leg of the duty period, respectively.

Sit time: The interval between two successive connected flight legs in a duty period.

Brief: The interval before commencing the first flight leg of the duty to prepare crews for the duty.

Debrief: The interval after completing the last flight leg of a duty period to prepare the next flight crews.

Crew base: The airport where crews are stationed and also the location where the duty period begins and ends.

Rest period: The interval between two successive duty periods (sometimes called overnight or layover rest, or overnight or layover connection). Within the rest period, crews are free from all sorts of duties.

Pairing: A sequence of one or more duty periods to be flown by a crew member, starting and ending at the same crew base, often with one or more rest periods interposed between them.

Flying time: The total hours of actual flying time.

Time away from base (TAFB): The elapsed time of a crew from departure to return to the same crew base, which includes the overnight rests between duty periods in a pairing, i.e. the total duration of a pairing.
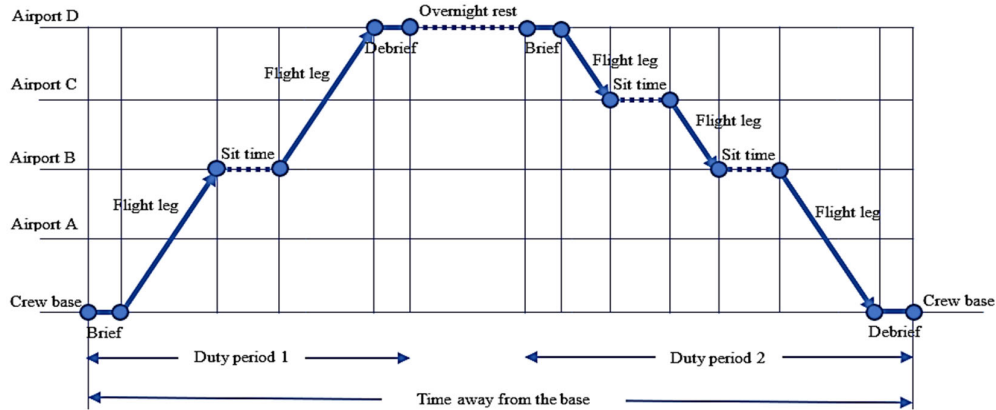

Fig. 1. A crew pairing example

A legal crew pairing must not violate local and international safety regulations and must comply with the company's agreed operational rules. The rules and regulations of the case-study budget airline stemming from the civil aviation authority of Thailand (CAAT), international civil aviation organisation (ICAO), and the airline's labour unions, are listed as follows.

- The CCPP in this study is prepared weekly since the flight schedule in the case-study airline is recurrent every week.
- Only one crew base, located at Don Muang international airport, is in operation.
- The policy of the case-study airline permits minimal deadheading to save crew costs.
- In a duty period, any two successive flight legs should be organised in such a way that the destination of the previous flight is the same as the departure of its direct consecutive flight.
- The brief and debrief times take 60 and 30 minutes, respectively.
- At most six flight legs could be assigned in a crew pairing.
- At least 30 minutes of the sitting time must be available between two consecutive flights.
- The maximum duration of a duty period each day depends on the duty start time and the number of flights in the duty period, as shown in Table 1.

Table 1
Maximum flight duty duration (h)

Duty start time	Maximum duty periods (h)						
	Number of flight legs						
	1	2	3	4	5	6	7 or more
05.00 – 05.59	13	13	12	12	12	11	11
06.00 – 12.59	13.30	13.30	13	12	12	12	11
13.00 – 14.59	13	13	12	12	12	11	11
15.00 – 04.59	12	12	11	11	11	10	10

- The accumulated duration of the duty periods responsible by each cockpit crew must not exceed the following values, i.e. ≤ 34 h within every 7 consecutive days, ≤ 110 h within every 28 consecutive days, and ≤ 1000 h within every 365 consecutive days.
- The minimum rest period given to a cockpit crew after completing each of his/her duty depends on the duty flight time, as shown in Table 2.

Table 2
Minimum rest duration (h)

Duty time (h)	Minimum rest duration (h)
< 8	≥ 8
8 – 10	≥ 10
10 – 12	≥ 12
12 – 14	≥ 14
14 – 16	≥ 16
16 – 20	≥ 24

Due to the cost-saving strategy of the budget airline, aircrew scheduling is conducted manually by an experienced planner using Excel. The objective of the CCPP is to determine a set of crew pairings with minimum costs that encompass all flight legs laid down in the timetable of the airline as well as satisfying all of the airline's restrictions. The input data of the CCPP is composed of a set of flight legs defined in the airline's timetable, fleet routing, costs of pairings, and legitimacy constraints. However, the identities, ranks, availability and combinations of different crew members are not considered in the CCPP. The following is the manual planning procedure of the case-study airline to realise appropriate cockpit crew pairings (CCPs).

- *Data preparation*: The flight legs, flight number, departure and destination airports, regions to fly of the country, flying distance (in nautical miles), departure and arrival times, block hours, and daily flight patterns in each week are updated and prepared by the planner.
- *Determination of the CCP*: The steps to manually plan CCPs with Excel are as follows.
    a) Sort all flights in ascending order of their departure times and initialise the pairing number. The first pairing number is O1, followed by O2, O3, …, etc.
    b) The first flight in any pairing must depart from Don Muang international airport (crew base).
    c) Check if the flight list is empty or not. If yes, the CCPP process is ended; otherwise, go to d).
    d) Select the flight (by strictly considering the regulations of CAAT and ICAO) with the earliest departure time from the list of flights whose airport of the next outbound flight matches the airport of the previous inbound flight. If feasible, add it into the current pairing, and update the list (by deleting the selected flight from the list); otherwise, go to step f).
    e) Check if the number of flight legs in the current pairing is more than six or not. If yes, stop adding more flights into the current pairing and go to step f); otherwise, go to step c).
    f) Increase the pairing number by 1 and go to step b).

In this paper, besides the CPC, which is broadly used in literature, other objectives deemed by the airline executives as equally important were also considered and optimised simultaneously. Let p be the pairing number, so $p \in P$ (total number of pairings). The cost structure of the pairing p, denoted as c_p, expressed in units of time, is computed as shown in Eqs. (1) and (2).

$$c_p = \max \{nd_p * m\_h_p, \sum_{d \in p} c_d, f_p * ta_p) \tag{1}$$

$$c_d = \max \{m\_h_d, fly_d, f_d * e_d\} \tag{2}$$

where d is the duty period number in the pairing p; nd_p is the number of duty periods in the pairing p; m_h_p is the minimum number of hours of the guaranteed pay for each pairing; c_d is the number of hours in the duty period d; ta_p is the time away from the crew base of the paring p; m_h_d is the minimum number of hours of the guaranteed pay for the duty period d; fly_d is the flying hours of the duty period d, e_d is the elapsed time of the duty period d; and $0 \le f_p$ and $f_d \le 1$ are the fractions of the elapsed times of the pairing and duty periods, respectively. The values of f_p and f_d can be varied according to the agreement of the airline. Note that we can simply convert the CPC (h) to any currency by multiplying c_p with a given currency factor. The objective to minimise the CPC is formulated as shown in Eq. (3).

$$f_1 = min \sum_{p \in P} c_p \tag{3}$$

The second objective is concerned with balancing the workload (i.e. TAFB) among different parings. It is desirable that the TAFB (ta_p) is uniformly distributed across different pairings. Let $\overline{ta_p}$ be the average TAFB of the CCP p. Minimising the mean absolute deviation of the TAFBs among different pairings is then formulated as shown in Eq. (4).

$$f_2 = \frac{1}{P} min \sum_{p \in P} |ta_p - \overline{ta_p}| \tag{4}$$

The third objective attempts to avoid adding repeated flight legs with the same start and end destinations in each crew paring. Apart from reducing the boredom of flying repeatedly on the same route, the cockpit crew can gain more experience and become familiar with diverse flight routes by utilising this objective. Let RFL_p be the number of repeated flight legs with the same start and end destinations in the crew paring p. Minimising the repeated flight legs is then given by Eq. (5).

$$f_3 = min \sum_{p \in P} RFL_p \tag{5}$$

The fourth objective attempts to balance the workload among different crew pairings, as measured in the form of nautical miles accumulated from the set of flights in the pairings. Let nm_p and $\overline{nm_p}$ be the total and average nautical miles of the pairing p, respectively. Minimising the mean absolute deviation of the nautical miles among different pairings is then given by Eq. (6).

$$f_4 = \frac{1}{P} min \sum_{p \in P} |nm_p - \overline{nm_p}| \tag{6}$$

The last objective attempted to minimise the number of crew pairings. The total number of CCPs (P) can be used to measure the minimum number of cockpit crews needed to be prepared by the airline (i.e. crew member capacity). A lower value of P means a higher utilisation of cockpit crews, resulting in the worthiness of the wages and benefits paid by the airline. This is formulated as shown in Eq. (7).

$$f_5 = \min P \tag{7}$$

4. Proposed algorithm

4.1 ANSDE III

The use of EAs has become an outstanding de facto approach nowadays in successfully dealing with complex real-world applications to achieve single- and multi-objectives. However, when the number of objectives is higher than three, traditional EAs are an ineffective means to find the Pareto front (PF) in which diverse non-dominated solutions (NDSs) are uniformly distributed and well-converged in high-dimensional space (2007). The reason is that the recombination and selection mechanisms of traditional EAs to guide the appropriate trajectory to the population to gear towards the true PF substantially lose their effectiveness since most of the offspring solutions are NDSs. To alleviate the above issues, EAs developed for optimising more than three objectives simultaneously, which emphasize the improvement of the diversity preservation mechanism rather than the Pareto domination-based principle in searching for NDSs, seems to be a more suitable tool. In this paper, two EAs (i.e. NSGA III and MOEA/D) (Aydemir-Karadag *et al*., 2013) and an EA which is effective in the offspring generation mechanism, i.e. the differential evolution (DE) algorithm, were integrated to induce the strength of each algorithm to form a single HA, namely the adaptive non-dominated sorting differential evolution III (ANSDE III) algorithm. The remarkable feature of NSGA III is its ability to search for NDSs that are evenly distributed. Meanwhile, MOEA/D and DE are capable of guiding the algorithms to converge towards good NDSs. The structure of ANSDE III is as follows.

A reference-point based NSGA II (namely NSGA III), the latest version of the non-dominated sorting GA (NSGA) collection developed by Deb and Jain (2014), adopts the diversity preservation concept in its evolutionary process. Multiple prespecified reference points scattered all over the whole search space are used in NSGA III for diversity preservation and directing multiple searches toward the goal positions along each direction. Due to its effectiveness, NSGA III has been widely accepted in optimising real-world MO optimisation problems (MOOPs).

The general framework of NSGA III is explained as follows (Ibrahim *et al*., 2016). At generation t, let P_t be the parent population of size N, Q_t be the offspring population of size N generated from P_t, S_t be the selected population, and E_t be the elite members.

1. *Reference point construction*: The reference points in NSGA III are used to warrant diversity in the offspring population. The determination of these points has to be done before commencing the evolutional process of NSGA III. The simplex-lattice design, one form of the mixture experimental design, is often adopted to systematically plot uniformly distributed normalised weight vectors on the hyperplane. If the problem has M objectives, as well as p separations with an equal space along with each objective and one intercept on each axis is required, we can calculate the entire number of reference points (H) by $H = C_p^{M+p-1}$.
2. *Initialisation*: The initial population (P_0) of size N is randomly created from the uniform distribution. Set $S_0 = \phi$, $E_0 = \phi$, and $t = 1$.
3. *Offspring generation*: Apply selected GOs to manipulate the crossover and mutation processes to create the offspring population (Q_t) from $P_t \cup E_t$.
4. *Population combination*: Find the combined population (R_t) of size $2N$ from $R_t = P_t \cup Q_t$.
5. *Non-dominated sorting*: Remove all repeated solutions in R_t. Apply a non-dominated sorting technique to rank R_t into different levels of non-dominated fronts, i.e. $F_1, F_2, ..., F_L$.
6. *Population selection*: Add non-dominate fronts of R_t to be the members of S_t in the new population P_{t+1}, front by front, starting from F_1, then $F_2, F_3, ...$ until $\|S_t\| \geq N$. Note $\|S_t\|$ means the number of members in the set.
    a. If the total members of S_t is equal to N (i.e. $\|S_t\| = N$), set $P_{t+1} = S_t$, and go to Step 7).
    b. If $\|S_t\| > N$, do the following.
        i. Find the first level (l) that causes $\|F_1 + F_2 + \cdots + F_l\| > N$. All population members from F_1 to F_{l-1} are contained within the members of P_{t+1}. The remaining members (RM) of P_{t+1} are selected from F_l.
        ii. Find $f_i^n(s)$ which is the normalised objective of each member $s \in S_t$ for each objective ($i = 1,2, ..., M$) using $f_i^n(s) = \{f_i(s) - z_i^{min}\}/\{z_i^{nad} - z_i^{min}\}$, where z_i^{min} is the minimum value of the objective i in S_t, and z^{nad} is the nadir point computed by $z^{nad} = max\{f_j(s)\}$.

      iii.   Associate each population member $s \in S_t$ with a reference point by selecting the reference point k of which its reference line is nearest to the population member $s \in S_t$ as the associate reference point.

      iv.   Find niche count ρ_k for each reference point k by counting the number of its associate population members $s \in S_t$, but not including the population members in F_l.

      v.   Identify the reference point k that has the lowest value of ρ_k. If more than one reference point is eligible, choose any one of them at random.

      vi.   If $\rho_k = 0$ and

            vii.1.   None of the members in F_l is associated with the reference point k, then no longer consider this reference point in this generation.

            vii.2.   Only one member in F_l is associated with the reference point k, then add that member in P_{t+1} and increase the value of ρ_k by one.

            vii.3.   More than one member in F_l is eligible, then select the one with the nearest perpendicular distance from the reference line, add that member in P_{t+1}, and increase the value of ρ_k by one.

      vii.   If $\rho_k > 0$ and

            vii.1.   Only one member in F_l is associated with the reference point k, then add that member in P_{t+1} and increase the value of ρ_k by one.

            vii.2.   More than one selectable member is available, then apply random selection, add that member in P_{t+1}, and increase the value of ρ_k by one.

      viii.   Repeat steps vi-vii for RM times (i.e. the remaining members of P_{t+1} are filled up).

7.   Update the elite members in E_t, set $S_t = \phi$ and $t = t + 1$.

8.   Repeat Steps 3 to 6 until the termination condition is reached.

The aforementioned steps in NSGA III are adopted as the main structure of ANSDE III. It appears that the offspring generation process of NSGA III uses GOs in the crossover and mutation executions to create high quality and diverse solutions. However, Chandrasekar and Ramana (2012) reported that the offspring generation mechanism of the DE algorithm was more effective than GA in searching for a global optimum. Moreover, Li and Zhang (2009) showed that in MOOPs, the replacement of the offspring generation process of GA by DE in the basic structure of the MOEA could enhance its performance substantially. Subsequently, the concept as such is implemented in ANSDE III and hence the offspring generation illustrated in Step 3 of NSGA III is replaced with the mechanism of DE. The use of DE was first pioneered by Storn and Price (1997) to optimise continuous complex optimisation problems and is an instance of EAs renowned for effectively solving various practical engineering problems and being simple to implement. Similar to GA, DE uses crossover and mutation mechanisms to create offspring. However, the reversed order of these two operators is applied in DE, i.e. first mutation followed by crossover. Moreover, a real number is used in DE to represent each gene in the solution's chromosome, instead of binary (0, 1) or positive integers as in GA. Although various mutation and crossover techniques have been proposed to create tentative offspring in DE (Price *et al.*, 2006), the DE/rand/1/bin strategy is used in ANSDE III due to its effectiveness, yet uncomplicatedness. The working process of this strategy is explained as follows. First, three solutions in the current population are chosen at random as parents. One of them is randomly assigned as a target vector (x_1), while the remainders (x_2 and x_3) are used for calculating the differential discrepancy. A mutant vector (v) is created from the mutation by adjusting the target vector x_1 with the weighted differential discrepancy, as shown in Eq. (8).

$$v = x_1 + F(x_2 - x_3) \tag{8}$$

where $F \in [0,1]$ is the weighted factor. The crossover starts immediately after the mutation is completed. During the crossover, the mutant vector (v) and the target vector (x_1) are combined at the randomly selected crossover positions j of the solution to form the trial vector x'. The trial vector at position j, namely x'_j, is created as shown in Eq. (9).

$$x'_j = \begin{cases} v_j, & \text{if } rand[0,1] < CR \\ x_j, & \text{otherwise} \end{cases} \tag{9}$$

where CR is the crossover probability. If the trial vector and the target vector are identical after the crossover process is completed, one position in the trial vector is randomly selected and the value of x_j is replaced by v_j. In addition, we will replace x_1 with x' if the trial solution dominates the target solution. The mating restriction and neighbouring solution updating mechanisms of the MOEA based on decomposition (MOEA/D) proposed by Zhang and Li (2007) are also utilised in ANSGA III to increase its convergence speed. For a reference point k, its associated neighbouring reference points NB_k (sized T) comprise T reference points having the shortest Euclidean distances from k. As a result, this process is added into Step 1 of the original NSGA III and renamed as "reference point and neighbourhood construction".

In Step 3 of the algorithm, the crossover process is limited to its neighbours only. After completing the first iteration of the algorithm, each reference point will have some population solution associated with it. Once the first solution (target vector) is randomly chosen from P_t, the remaining solutions that will join the crossover process are the associated solutions of the neighbouring reference points of the reference point associated with the first solution. Since the solutions involved in the

crossover process come from the current best neighbouring solutions of the first selected solution, we hoped that a good offspring would be created. The following steps proceed after the offspring solution is created. If the target solution dominates the offspring solution, the offspring solution is discarded from further consideration. If the offspring dominates the target solution, the target solution is replaced with the offspring and a reference point for the offspring solution is found. If the offspring solution and the target solution are non-dominated with each other, both of them are kept and a reference point to the offspring solution is assigned. To further enhance the global search of the algorithm, an adaptive mechanism is also embedded in ANSDE III. The status of the elite members in the external memory evolved over time is monitored to identify if the parameters of the algorithm need to be adjusted. In every g iterations, the alteration of the elite members is probed. Note that the value of g is set to 10 in this research. Three main parameters of the algorithm that could pull the current solutions off from the local optima are the weighted factor (F), crossover probability (CR), and the neighbourhood size (Nb). If the elite solutions remain unchanged for the past g iterations, the values of F, CR, and T will be changed; otherwise, these parameters are kept at their original values. The value of F affects the resemblance between the target solution and the mutant solution in that a higher value of F could increase their discrepancies. The value of CR controls the similarity between the trial and target solutions. If the value of CR is low, the trial and target solutions will not be very different. In addition, the value of Nb is used to control the selectable solution members to join the crossover process. A high value of Nb allows the solution members that are quite distant from the target solution to participate in the crossover process. As a result, the values of F, CR, and Nb should be set at high if a high diversity in the solutions is required to be created (i.e. the elite solutions have not changed during the last g generations). Otherwise, these parameters are maintained at low if some changes are detected in the elite members. This process is added at the end of Step 7 (i.e. parameter adaptation mechanism).

4.2 Solution encoding

The position-based priority scheme is applied to represent the solution string (chromosome) of the CCPP. The length of the string is equal to the number of flight legs in the problem. The positions from left to right in the string indicate the assignment priority of the flight legs (i.e. the highest priority is given to the flight leg that is assigned in the left-most position). For example, if eight flight legs are considered as shown in Table 3, a possible solution string can be A1, A2, A3, A4, A5, A6, A7, and A8.

Table 3
Example of flight legs

Flight No.	Code	From	To	Departure	Arrival	Flight Time
A1	XX8714	DMK	CEI	7.20	8.40	1.20
A2	XX8715	CEI	DMK	9.20	10.35	1.15
A3	XX8308	DMK	CNX	11.10	12.20	1.10
A4	XX8312	CNX	DMK	12.40	13.50	1.10
A5	XX8622	CNX	UTH	14.10	15.15	1.05
A6	XX8625	UTH	CNX	19.10	20.25	1.15
A7	XX8302	DMK	CNX	7.00	8.10	1.10
A8	XX8303	CNX	DMK	9.10	10.10	1.10

4.3 Solution decoding

The output derived after decoding the solution string is the physical pairings. The first-fit rule was applied to construct pairings. Since the position-based priority scheme was implemented in the encoding process, we had to sweep from left to right of the solution string and the first flight leg that did not violate the regulations discussed in Section 3 was selected. The selected flight leg was assigned to the pairing and was scheduled in the flight timetable. The first-fit assignment process was then repeated until no more flight legs in the solution string were left. According to the solution string exemplified in the encoding process (i.e. A1, A2, A3, A4, A5, A6, A7, and A8), the pairings and their corresponding timetables are shown in

Fig. 2 Solution string encoding

5. Experimental design

Three well-known algorithms developed for solving MOOPs were employed to compare their relative performances with ANSDE III. These were MOEA/D (Zhang and Li, 2007), NSGA III (Deb and Jain, 2014), and MODE (Xue *et al*., 2003). These algorithms were evaluated using the same testbeds through various instance problems, and the details of the testbeds and parameter settings for each algorithm are discussed as follows.

5.1 Instance problems

To conduct a rigorous performance test of ANSDE III, various problem instances were used as the testbeds. The instances ranged from small (S) to medium (M) and large (L), based on the number of flight legs. Some data, particularly in the large instances, were obtained from the flight schedule of the case-study airline during the low and high season for travelling in Thailand. Also, these data are based on the Boing 737-800 fleet due to its highest capacity and number of aircraft. Although the small and medium instances are hypothetical, these data were generated at random based on truncated real data from the case-study airline. All instances used Don Muang international airport as the crew base. The details of the testbeds are shown in Table 4.
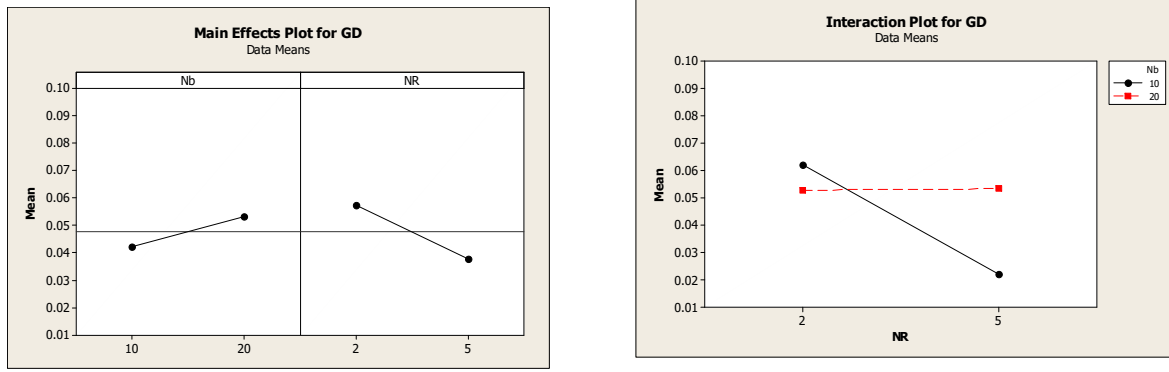
Table 4
Problem instances

| Problem size	Instance	Flight legs
Small	S1	20
	S2	40
	S3	60
	S4	80
Medium	M1	100
	M2	120
	M3	150
	M4	152
Large	L1	200
	L2	282
	L3	298
	L4	320

5.2 Parameter setting

The algorithms that competed in finding non-dominated solutions (NDSs) for the MOCCPP comprised ANSDE III, NSGA III, MOEA/D, and MODE. To ensure fair competition, the key parameters, which are the drivers of each algorithm, were identified. The initial values of these parameters were obtained from the literature. The design of experiments was based on the factorial design concept (Montgomery, 2017) and was used to determine the best parameter settings suitable for each algorithm. As mentioned by Jiang *et al.* (2014) and Lu *et al.* (2019), the generational distance (GD), invert generational distance (IGD) and $Spread$ (sometimes called the spacing metric (SM) or symbolised as △) were the three major metrics for assessing the relative efficacy of different Pareto fronts (PFs) obtained by multi-objective evolutionary algorithms (NB: the descriptions of these and other metrics used in this paper will be given in Section 5.3). We prioritised the GD first, then IGD and $Spread$ last since a better PF should lay close toward or even overlap with the true-Pareto front (TPF) and also uniformly distributed. Given that the levels of each factor (parameter) to be tested for each algorithm were two. As a result, a 2^k factorial design was applied in determining suitable values of the algorithm's parameters. A hierarchical Pareto performance metric satisfaction approach was taken in this matter. First, the results from the analysis of variance (ANOVA) on the GD were considered. If a significant difference was noticeable on main effects or interactions (any p-value ≤ 0.05), the parameters were then set to the treatment combination's values that minimise the GD. Otherwise (all p-values > 0.05), move from the GD to IGD and apply the same approach to the IGD. The parameter settings would be based on the treatment combinations that result in the lowest IGD if any statistically significant effect was revealed. If we still cannot observe any significant effect on the IGD from ANOVA (all p-values > 0.05), continue to apply this approach to the $Spread$. After that, finally, if no significant effect was detected yet (all p-values > 0.05), the parameters were then set to those parameter's levels that made the GD minimise. For example, suppose the parameters (i.e. Nb and NR) of MOEA/D for the small-sized problem (S2) needed to be established. Two levels of the factors Nb and NR were experimented with, i.e. (10, 20) and (2, 5), respectively. First, the 2^2 factorial experiments were conducted to determine whether or not these settings in combination had any effects on the GD. The results of ANOVA revealed that the main effects and two-way interactions were not significant (all p-values > 0.05) at a 95% confidence level ($\alpha = 0.05$). Figure 3 (a) depicted the main effect and interaction plots for the GD. As a result, the next step was to conduct the experiments on the IGD. The ANOVA's result summarised that the main effect of Nb had a significant effect on the IGD (p-value ≤ 0.05). Since the lowest value of the IGD was desirable, the interaction plot (Figure 3 (b)) was observed to determine the values of the Nb and NR parameters as such, i.e. $Nb = 10$ and $NR = 5$. Table 5 summarised the appropriate parameters for each algorithm which were the results of applying the hierarchical Pareto performance metric satisfaction approach based on the 2^k factorial design analysis. The number of populations was 131 based on a simplex lattice

design (Deb and Jain, 2014) and the number of generations to be evolved was 2,000 for all problems and algorithms. All algorithms were coded in MathLab R2019a under a personal computer equipped with Intel(R) Core™ i7-8550U CPU @ 1.8GHz 8.0 GB RAM, running under the Microsoft Windows 10 Pro 64-bit operating system.

(a) Main effect and interaction plot for the GD metric
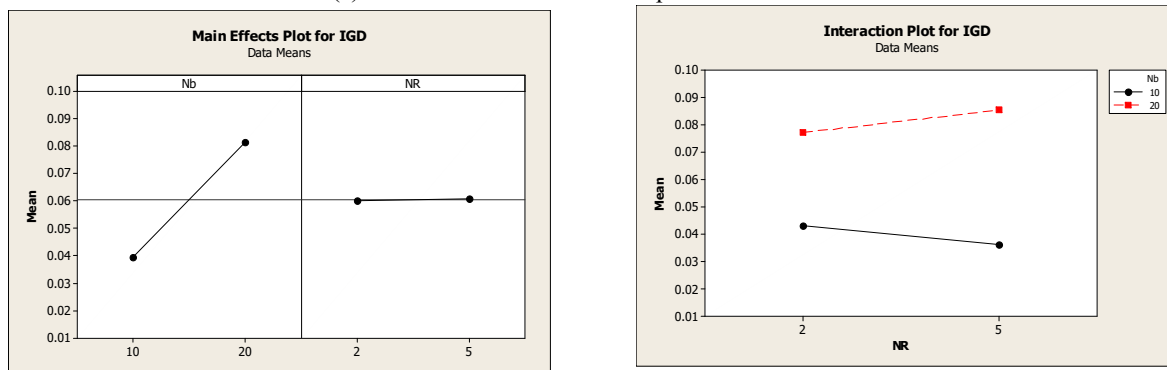
(b) Main effect and interaction plot for the IGD metric

Fig. 3. 2^2 factorial design of experiments to determine the MOEA/D's parameters

Table 5
Parameter settings for each algorithm

No. of Flights	Problem Size	Set	MOEA/D		NSGA III		MODE		ANSDE III		
			Nb	NR	Pc	Pm	F	CR	F	CR	Nb
20		S1	10	5	0.9	0.4	1	0.9	0.8, 1	0.7, 0.9	10
40		S2	10	5	0.6	0.1	1	0.9	0.8, 1	0.7, 0.9	10
60	Small	S3	20	2	0.9	0.1	1	0.9	0.8, 1	0.7, 0.9	10
80		S4	10	2	0.6	0.1	1	0.9	0.8, 1	0.7, 0.9	10
100		M1	10	5	0.6	0.1	1	0.9	0.8, 1	0.7, 0.9	10
120		M2	10	2	0.6	0.1	1	0.9	0.8, 1	0.7, 0.9	10
140	Medium	M3	10	2	0.6	0.4	1	0.9	0.8, 1	0.7, 0.9	10
152		M4	10	2	0.9	0.1	1	0.9	0.8, 1	0.7, 0.9	10
200		L1	10	2	0.9	0.1	1	0.9	0.8, 1	0.7, 0.9	10
282	Large	L2	10	2	0.6	0.4	1	0.9	0.8, 1	0.7, 0.9	10
298		L3	10	2	0.6	0.4	1	0.9	0.8, 1	0.7, 0.9	10
320		L4	10	2	0.6	0.4	1	0.9	0.8, 1	0.7, 0.9	10

(Note: Nb is the number of weight vectors in the neighbourhood, NR is the maximum number of solutions replaced by each offspring, Pc is the probability of crossover, Pm is the probability of mutation, F is the scaling factor, CR is the crossover rate)

5.3 Performance measurements

Various Pareto-based metrics were recommended in the literature to compare the relative performance of multi-objective optimisation algorithms in the Pareto sense. Some of them were identical since they gauged the same things but named differently. Therefore, to demonstrate their linkages particularly with our previous research (Chutima and Kirdphoksap, 2019), notes on synonyms were added as appropriate. In this paper, the metrics to measure the performance of ANSDE III against NSGA III, MOEA/D, and MODE in optimising the MOCCPP include the generational distance (GD), Convergence ($Conv$), inverted generational distance (IGD), ratio of NDS located on the TPF based on its own PF (R_{NDS1}), quality metric (QM;

sometimes named ratio of NDSs located on the TPF based on the TPF (R_{NDS2})), *Spread* (sometimes symbolised △ or named spacing metric (*SM*)), number of Pareto solutions (*NPS*; sometimes named number of generated non-dominated solutions (N_{NDS})), best frontier members (*BFM*), mean ideal distance (*MID*), rate of achievement to the objectives simultaneously (*RAS*), and diversification metric (*DM*). These metrics are widely accepted as the standard performance measurements of multi-objective EAs, especially *GD* and *IGD*, in optimising MO problems (Jiang *et al*., 2014 and Lu *et al*., 2019). The detailed formulations of these metrics are explained in (Chutima and Olarnviwatchai, 2018; Chutima and Kirdphoksap, 2019; Zade *et al*., 2014; Asefi *et al*., 2014). In this research, the TPF is approximated by merging all non-dominated solutions (NDSs) obtained from all algorithms, applying non-dominated sorting to all of them, and then selecting all NDSs on the first front as the TPF.

6. Results

To signify the statistical property of some input parameters used in the contested algorithms, which are based on random number generators, all algorithms were run with three replicates. The best values (minimum) of the objectives obtained (*BFM*) from different algorithms were collected. Since some results are sensitive information of the case-study airline and also under the terms of the legal contractual agreement, the actual values cannot be disclosed to the public. However, their comparative data were allowed to present in a sense of proportion to each other just for academic purposes only. As a result, the data from ANSDE III were used as a base for the relative performance comparisons and so they were set to be equal to 1. The values of the objectives obtained from the other algorithms compared with those from ANSDE III were derived from dividing the actual values of those algorithms by the actual values of ANSDE III. From Table 6, it is clear that ANSDE III always achieved the best *BFM* metric (i.e. for the best values of the objectives, the lower the better). Meanwhile, MOEA/D and NSGA III could also reach the best value in several cases. However, the number of cases that MODE gave the best value was very few compared with the other algorithms and was restricted to a few small-sized problems only.

Table 6
Best objectives (*BFM*) found by the four different algorithms

Algorithm	Problem											
	S1	S2	S3	S4	M1	M2	M3	M4	L1	L2	L3	L4
f1: Minimize pairing cost												
MOEA/D	1.00	1.00	1.00	2.32	1.82	1.44	1.43	1.93	1.06	1.87	1.92	2.00
NSGA III	1.00	1.00	1.00	1.00	1.60	1.30	1.35	1.16	1.04	1.45	1.45	1.45
MODE	1.00	1.00	2.14	2.93	2.51	2.41	2.96	2.16	2.36	2.44	2.51	2.63
ANSDE III	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
f2: Balance workload (TAFB)												
MOEA/D	1.00	1.00	1.00	1.00	1.00	1.00	1.15	1.00	1.01	1.13	1.23	1.23
NSGA III	1.00	1.00	1.00	1.00	1.01	1.01	1.16	1.14	1.01	1.08	1.09	1.11
MODE	1.00	1.00	1.01	1.24	1.09	1.01	1.32	1.24	1.11	1.24	1.25	1.26
ANSDE III	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
f3: Minimize repeated flight patterns												
MOEA/D	1.00	1.00	1.00	1.00	1.00	1.00	1.15	1.17	1.34	1.05	1.00	1.03
NSGA III	1.00	1.00	1.00	1.00	1.00	1.00	1.20	1.19	1.17	1.03	1.02	1.02
MODE	1.00	1.00	1.53	1.52	1.52	1.34	1.34	1.34	1.51	1.07	1.05	1.05
ANSDE III	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
f4: Balance workload (nautical miles)												
MOEA/D	1.00	1.00	1.00	1.00	1.09	1.17	1.17	1.14	1.10	1.13	1.23	1.24
NSGA III	1.00	1.00	1.00	1.00	1.09	1.17	1.17	1.14	1.10	1.08	1.08	1.08
MODE	1.00	1.00	1.64	1.50	1.39	1.17	1.33	1.14	1.21	1.24	1.25	1.36
ANSDE III	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
f5: Minimize the number of pairings												
MOEA/D	1.00	1.00	1.00	1.00	1.00	1.00	1.09	1.17	1.20	1.09	1.09	1.07
NSGA III	1.00	1.00	1.00	1.00	1.00	1.00	1.09	1.17	1.13	1.05	1.04	1.04
MODE	1.00	1.00	1.30	1.50	1.29	1.22	1.18	1.18	1.27	1.14	1.13	1.11
ANSDE III	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

To evaluate the characteristics of the PFs obtained from different algorithms in comparison with the TPF, several metrics were employed in this study, i.e. *GD*, *Conv*, *IGD*, R_{NDS1}, *QM* (or R_{NDS2}), *Spread* (or *SM*), *NPS* (or N_{NDS}), *MID*, *RAS* and *DM*. Liu *et al*. (2020) proposed four key categories to measure the efficacy of metaheuristics in optimising discrete multi-objective problems in their recent state-of-the-art review, i.e. (1) quantitative, (2) convergence-related, (3) distribution-related and (4) comprehensive-related metrics. Consequently, the metrics used in this research were chosen by at least one member from each category. The selected metrics were among those frequently used in literature and fallen into Liu *et al*. (2020)'s categories as follows: *NPS* was the representative of the quantitative metric; *GD*, *Conv*, R_{NDS1}, *QM*, *MID* and *RAS* were the representatives of the convergence-related metric; *Spread* and *DM* were the representatives of the distribution-related metric; *IGD* was the representative of the comprehensive-related metric.

A short description of each metric can be explained as follows. As a member of the quantitative measurement, the *NPS* counts the number of NDSs located in the first front (i.e. Pareto front, PF) of the given algorithm. Regarding the convergent-related measurement, the *GD* and *Conv* are similar in terms of denoting the closeness between the NDSs in the PF of the given

algorithm and those in the TPF. However, these two metrics differ in that the GD uses the shortest distance between each NDS obtained from the given algorithm and the NDS in the TPF in its computation; whereas the $Conv$ uses the average distance of each NDS obtained from the given algorithm and those NDSs in the TPF. The R_{NDS1} and QM demonstrate the ratio between the number of the NDSs of the given algorithm that overlaps with the NDSs of the TPF and the number of its NDSs (R_{NDS1}), and the number of the NDSs of the TPF (QM), respectively. The MID and RAS reveal the nearness of the obtained NDSs to the ideal points (i.e. best value of each objective). While the MID measures in terms of the root mean of different square values, the RAS reports a mean absolute of different values. Under the distribution measurement, the $Spread$ indicates the dispersal of the NDSs on the PF of the given algorithm in a mean absolute different value. Meanwhile, the DM demonstrates the boundary between maximum and minimum values (range) of the NDSs in each objective obtained by the algorithm. Finally, the IGD, which is a member of the comprehensive measurement, is capable of demonstrating both the diversity and convergence characteristics of the NDSs in the PF of the given algorithm. Note that the lower values of GD, $Conv$, IGD, $Spread$, MID and RAS are superior; whereas, the higher values of NPS, R_{NDS1} and DM are desirable.

Table 7
Pareto-based measurements obtained by the different four algorithms

Algorithm	Problems																	
	S1	S2	S3	S4	Avg	Rank	M1	M2	M3	M4	Avg	Rank	L1	L2	L3	L4	Avg	Rank
Quantitative Measurement																		
Number of Pareto solutions (NPS)																		
MOEA/D	5	6	7	8	6.5	3	10	12	14	14	12.5	3	21	23	25	27	24.0	3
NSGA III	6	7	8	10	7.7	2	12	13	15	16	14.0	2	22	24	26	28	25.0	2
MODE	4	5	6	7	5.5	4	9	10	11	11	10.2	4	20	22	24	26	23.0	4
ANSGA III	6	8	10	12	9.0	1	13	15	17	18	15.7	1	23	25	28	30	26.5	1
Convergent-Related Measurement																		
Generational Distance (GD)																		
MOEA/D	0.31	0.05	0.04	0.11	0.13	3	0.12	0.27	0.18	0.31	0.22	3	0.04	0.03	0.04	0.03	0.03	3
NSGA III	0.11	0.05	0.04	0.06	0.06	2	0.09	0.21	0.05	0.17	0.13	2	0.03	0.02	0.03	0.03	0.03	2
MODE	0.40	0.07	0.06	0.13	0.17	4	0.28	0.37	0.19	0.64	0.37	4	0.08	0.03	0.05	0.04	0.05	4
ANSGA III	0.04	0.02	0.03	0.04	0.04	1	0.05	0.05	0.03	0.07	0.05	1	0.03	0.01	0.02	0.01	0.02	1
R_{NDS1}																		
MOEA/D	0.34	0.16	0.19	0.13	0.20	3	0.14	0.11	0.17	0.12	0.13	3	0.17	0.16	0.44	0.37	0.29	3
NSGA III	0.35	0.18	0.24	0.17	0.24	2	0.15	0.11	0.18	0.13	0.14	2	0.24	0.17	0.57	0.56	0.38	2
MODE	0.33	0.11	0.15	0.11	0.18	4	0.11	0.10	0.10	0.09	0.10	4	0.12	0.11	0.31	0.34	0.22	4
ANSGA III	0.37	0.28	0.33	0.30	0.32	1	0.26	0.35	0.31	0.30	0.30	1	0.45	0.21	0.59	0.71	0.49	1
Quality metric (QM)																		
MOEA/D	0.25	0.08	0.08	0.08	0.13	3	0.08	0.08	0.08	0.08	0.08	2	0.13	0.12	0.32	0.27	0.21	3
NSGA III	0.30	0.08	0.08	0.08	0.14	2	0.08	0.08	0.08	0.08	0.08	2	0.15	0.13	0.42	0.41	0.28	2
MODE	0.13	0.08	0.08	0.08	0.10	4	0.08	0.08	0.08	0.08	0.08	2	0.08	0.08	0.23	0.25	0.16	4
ANSGA III	0.33	0.17	0.17	0.17	0.21	1	0.17	0.17	0.17	0.17	0.17	1	0.29	0.15	0.44	0.52	0.35	1
Convergence (Conv)																		
MOEA/D	0.41	0.39	0.45	0.52	0.44	3	0.48	0.43	0.34	0.25	0.37	3	0.12	0.28	0.19	0.32	0.23	3
NSGA III	0.22	0.25	0.31	0.34	0.28	2	0.45	0.41	0.30	0.28	0.36	2	0.09	0.21	0.05	0.17	0.13	2
MODE	0.53	0.64	0.78	0.82	0.69	4	0.50	0.42	0.35	0.35	0.41	4	0.29	0.38	0.20	0.67	0.39	4
ANSGAIII	0.11	0.22	0.24	0..33	0.23	1	0.35	0.32	0.29	0.24	0.30	1	0.05	0.05	0.03	0.07	0.05	1
Mean ideal distance (MID)																		
MOEA/D	0.31	0.04	0.09	0.10	0.13	3	0.21	0.36	0.14	0.18	0.22	3	0.05	0.04	0.04	0.03	0.04	3
NSGA III	0.11	0.05	0.10	0.13	0.09	2	0.12	0.15	0.16	0.17	0.15	2	0.04	0.03	0.02	0.01	0.03	2
MODE	0.56	0.13	0.15	0.24	0.27	4	0.26	0.45	0.29	0.74	0.43	4	0.06	0.05	0.04	0.04	0.05	4
ANSGA III	0.06	0.04	0.01	0.10	0.03	1	0.03	0.06	0.02	0.01	0.02	1	0.03	0.02	0.02	0.01	0.02	1
Rate of achievement to the objectives simultaneously (RAS)																		
MOEA/D	0.09	0.06	0.07	0.15	0.11	3	0.19	0.31	0.12	0.16	0.19	3	0.01	0.02	0.04	0.03	0.03	3
NSGA III	0.23	0.01	0.11	0.11	0.06	2	0.11	0.13	0.14	0.15	0.14	2	0.02	0.01	0.03	0.03	0.02	2
MODE	0.56	0.35	0.18	0.21	0.31	4	0.23	0.40	0.25	0.65	0.39	4	0.04	0.03	0.05	0.04	0.04	4
ANSGA III	0.06	0.02	0.03	0.06	0.05	1	0.02	0.04	0.03	0.05	0.03	1	0.01	0.01	0.01	0.01	0.01	1
Distribution Measurement																		
Spread																		
MOEA/D	0.79	0.14	0.05	0.10	0.27	3	0.10	0.09	0.30	0.09	0.14	3	0.44	0.31	0.46	0.37	0.40	3
NSGA III	0.75	0.06	0.04	0.08	0.23	2	0.08	0.08	0.03	0.07	0.06	2	0.40	0.19	0.37	0.32	0.32	2
MODE	0.94	0.99	0.08	0.12	0.53	4	0.10	0.29	0.99	0.24	0.41	4	0.55	0.39	0.56	0.45	0.49	4
ANSGA III	0.57	0.02	0.03	0.03	0.16	1	0.06	0.02	0.01	0.05	0.03	1	0.10	0.13	0.19	0.16	0.15	1
Diversification metric (DM)																		
MOEA/D	0.35	0.41	0.45	0.49	0.40	3	0.49	0.55	0.62	0.68	0.59	3	0.71	0.75	0.82	0.87	0.79	3
NSGA III	0.38	0.46	0.51	0.55	0.45	2	0.52	0.59	0.67	0.71	0.62	2	0.76	0.78	0.81	0.89	0.81	2
MODE	0.27	0.32	0.39	0.45	0.33	4	0.48	0.52	0.55	0.58	0.53	4	0.65	0.68	0.75	0.79	0.72	4
ANSGA III	0.39	0.48	0.54	0.57	0.47	1	0.65	0.69	0.72	0.75	0.70	1	0.79	0.84	0.89	0.95	0.87	1
Comprehensive Measurement																		
Inverted Generational Distance (IGD)																		
MOEA/D	0.34	0.06	0.12	0.16	0.17	3	0.26	0.45	0.20	0.30	0.30	2	0.03	0.03	0.05	0.04	0.04	3
NSGA III	0.12	0.06	0.11	0.15	0.11	2	0.14	0.16	1.18	0.20	0.42	3	0.02	0.02	0.04	0.03	0.03	2
MODE	0.58	0.13	0.15	0.25	0.28	4	0.27	0.47	0.30	0.76	0.45	4	0.05	0.04	0.06	0.05	0.05	4
ANSGA III	0.06	0.02	0.04	0.07	0.05	1	0.07	0.13	0.08	0.14	0.10	1	0.02	0.01	0.02	0.02	0.02	1

Table 7 shows the relative performances of different algorithms on small-, medium- and large-sized problems (i.e. S, M and L) under four key measurement categories. For the quantitative measurement category, in which the NPS represented the group, the values obtained from ANSDE III were always better than the others. This indicated that the greater number of NDSs was created by ANSDE III in its PF. However, this metric could be deceptive since it would be meaningless if most of its NDSs are not located close to the TPF. Hence, this metric needs to be construed in conjunction with other more important measurements in rationally describing the results. Six metrics were used as representatives in the convergent-related category, i.e. GD, R_{NDS1}, QM, $Conv$, MID and RAS. It was observed that ANSDE III achieved the lowest values of the GD, $Conv$, MID and RAS; meanwhile, having the highest values of the R_{NDS1} and QM. If we take the results from the metrics in this category in combination with the previous category (NPS) and interpret them all together, it was obvious that not only ANSDE III could reach a higher number of NDSs but also their NDSs were positioned closer to the TPF than the others. Based on the $Spread$ and DM (distribution category) which indicate how uniform the NDSs obtained from a given algorithm

is distributed, ANSDE III attained lower values of $Spread$, and higher values of DM than others. This means the NDSs of ANSDE III were more uniformly distributed with a greater boundary range between maximum and minimum values in each objective than the others. Moreover, under the comprehensive category, the IGD represented a member of this group. This metric reflected the effectiveness in terms of the convergence and spread of the PF simultaneously. The result showed that ANSDE III outperformed the others since its IGD was lower. To demonstrate the overall performance of different algorithms based on different problem sizes (i.e. S, M and L), the resultant metrics of different algorithms were averaged and ranked accordingly. For different problem instances, the values of GD, $Conv$, IGD, $Spread$, MID and RAS of ANSDE III were always the lowest and its NPS, R_{NDS1}, QM and DM values were always the highest compared with the other three algorithms. This pattern of results was also shown in the average values obtained from ANSDE III regardless of the problem size (S, M, or L). In terms of comparative performances, their rankings from best to worst were as follows: ANSDE III > NSGA III > MOEA/D > MODE (note: the symbol " > " denotes better than). Since the number of outputs obtained from all the compared algorithms was not high enough to validly assume that their distributions were normal, the non-parametric Kruskal-Wallis test of the equality of medians for four algorithms (i.e. MOEA/D, NSGA III, MODE, and ANSGA III) under different problem sizes and performance measures was performed. The hypotheses stated as follows: (H_0) the population medians are all equal, as against (H_1) not all medians are equal. The results testified that there were differences between the population medians (p-value ≤ 0.05), and ANSDE III always ranked first among the matched algorithms. Moreover, the results from the two-sample Wilcoxon rank-sum test of the equality of two population medians signified that ANSDE III commonly outperformed the other algorithms.

7. Conclusion

The MOCCPP was addressed in this paper. Apart from the cost-concern objective, four additional objectives were also optimised simultaneously, i.e. balance workload (TAFB) among different parings, minimise repeated flight patterns, balance the workload (nautical miles) among different pairings, and minimise the number of pairings. Since the problem belongs to the NP-hard class of combinatorial optimisation problems, ANSDE III, an EA that is a hybrid of the NSGA III, MOEA/D, and DE algorithms, plus a simple adaptive mechanism was proposed. The performance of ANSDE III was compared against its original algorithms (NSGA III, MOEA/D, and MODE) under convergence-related, spread-related, and mixed metrics. In all cases, ANSDE III performed better than the other algorithms, in terms of achieving both the best values of different objectives and the obtained PFs.

Acknowledgement

References

Aggarwal, D., Saxena, D. K., Back, T., & Emmerich, M. (2020). Real-World Airline Crew Pairing Optimization: Customized Genetic Algorithm versus Column Generation Method. *arXiv preprint arXiv:2003.03792*.

AhmadBeygi, S., Cohn, A., & Weir, M. (2009). An integer programming approach to generating airline crew pairings. *Computers & Operations Research*, *36*(4), 1284-1298.

Anbil, R., Forrest, J. J., & Pulleyblank, W. R. (1998). Column generation and the airline crew pairing problem. *Documenta Mathematica*, *3*(1), 677-686.

Anbil, R., Gelman, E., Patty, B., & Tanga, R. (1991). Recent advances in crew-pairing optimization at American Airlines. *Interfaces*, *21*(1), 62-74.

Asefi, H., Jolai, F., Rabiee, M., & Araghi, M. T. (2014). A hybrid NSGA-II and VNS for solving a bi-objective no-wait flexible flowshop scheduling problem. *The International Journal of Advanced Manufacturing Technology*, *75*(5-8), 1017-1033.

Aydemir-Karadag, A., Dengiz, B., & Bolat, A. (2013). Crew pairing optimization based on hybrid approaches. *Computers & Industrial Engineering*, *65*(1), 87-96.

Barnhart, C., Hatay, L., & Johnson, E. L. (1995). Deadhead selection for the long-haul crew pairing problem. *Operations Research*, *43*(3), 491-499.

Barnhart, C., & Talluri, K. T. (1997). Airline operations research. Design and operation of civil and environmental engineering systems. Ed. Revelle, C. and McGarity, A. E., 435-469. New York, Wiley.

Chandrasekar, K., & Ramana, N. (2012). Performance comparison of GA, DE, PSO and SA approaches in enhancement of total transfer capability using FACTS devices. *Journal of Electrical Engineering and Technology*, 7(4), 493-500.

Chutima, P., & Kirdphoksap, T. (2019). Solving Many-Objective Car Sequencing Problems on Two-Sided Assembly Lines Using an Adaptive Differential Evolutionary Algorithm. *Engineering Journal*, *23*(4), 121-156.

Chutima, P., & Olarnviwatchai, S. (2018). A multi-objective car sequencing problem on two-sided assembly lines. *Journal of Intelligent Manufacturing*, *29*(7), 1617-1636.

Deb, K., & Jain, H. (2013). An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, *18*(4), 577-601.

Demirel, N. Ç., & Deveci, M. (2017). Novel search space updating heuristics-based genetic algorithm for optimizing medium-scale airline crew pairing problems. *International Journal of Computational Intelligence Systems*, *10*(1), 1082-1101.

Desaulniers, G., Desrosiers, J., Dumas, Y., Marc, S., Rioux, B., Solomon, M. M., & Soumis, F. (1997). Crew pairing at Airfrance. *European Journal of Operational Research*, *97*(2), 245-259.

Deveci, M., & Demirel, N. C. (2018). Evolutionary algorithms for solving the airline crew pairing problem. *Computers & Industrial Engineering*, *115*, 389-406.

Gopalakrishnan, B., & Johnson, E. L. (2005). Airline crew scheduling: state-of-the-art. *Annals of Operations Research*, *140*(1), 305-337.

Graves, G. W., McBride, R. D., Gershkoff, I., Anderson, D., & Mahidhara, D. (1993). Flight crew scheduling. *Management Science*, *39*(6), 736-745.

Ibrahim, A., Rahnamayan, S., Martin, M. V., & Deb, K. (2016, July). EliteNSGA-III: An improved evolutionary many-objective optimization algorithm. In 2016 *IEEE Congress on Evolutionary Computation (CEC)* (pp. 973-982). IEEE.

Jiang, S., Ong, Y. S., Zhang, J., & Feng, L. (2014). Consistencies and contradictions of performance metrics in multiobjective optimization. *IEEE Transactions on Cybernetics*, *44*(12), 2391-2404.

Klabjan, D., Johnson, E. L., Nemhauser, G. L., Gelman, E., & Ramaswamy, S. (2001). Solving large airline crew scheduling problems: Random pairing generation and strong branching. *Computational Optimization and Applications 20*(1), 73-91.

Lavoie, S., Minoux, M., & Odier, E. (1988). A new approach for crew pairing problems by column generation with an application to air transportation. *European Journal of Operational Research*, *35*(1), 45-58.

Li, H., & Zhang, Q. (2009). Multi-objective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II. *IEEE Transactions on Evolutionary Computation*, *13*(2), 284-302.

Liu, T. K., Chen, C. H., Chou, J. H., Chen, S. H., & Chou, T. Y. (2009, August). Application of multiobjective genetic algorithms for optimizing aircraft crew pairing problems. In *2009 ICCAS-SICE* (pp. 3748-3753). IEEE.

Liu, Q., Li, X., Liu, H., & Guo, Z. (2020). Multi-objective metaheuristics for discrete optimization problems: A review of the state-of-the-art. *Applied Soft Computing*, 106382.

Lu, C., Gao, L., Pan, Q., Li, X., & Zheng, J. (2019). A multi-objective cellular grey wolf optimizer for hybrid flowshop scheduling problem considering noise pollution. *Applied Soft Computing*, *75*, 728-749.

Montgomery, D. C. (2017). *Design and analysis of experiments*. John Wiley & Sons.

Price, K., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media.

Reisi, N. M., & Moslehi, G. (2013). Cockpit crew pairing problem in airline scheduling: Shortest path with resources constraints approach. *International Journal of Industrial Engineering and Production Researc*h, *24*(4), 259-268.

Souai, N., & Teghem, J. (2009). Genetic algorithm-based approach for the integrated airline crew-pairing and rostering problem. *European Journal of Operational Research, 199*(3), 674-683.

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, *11*(4), 341-359.

Xue, F., Sanderson, A. C., & Graves, R. J. (2003, December). Pareto-based multi-objective differential evolution. In *The 2003 Congress on Evolutionary Computation, 2003. CEC'03.* (Vol. 2, pp. 862-869). IEEE.

Zade, A. E., Sadegheih, A., & Lotfi, M. M. (2014). A modified NSGA-II solution for a new multi-objective hub maximal covering problem under uncertain shipments. *Journal of Industrial Engineering International*, *10*(4), 185-197.

Zeren, B., & Özkol, İ. (2012). An improved genetic algorithm for crew pairing optimization. *Journal of Intelligent Learning Systems and Applications*, *4*(1), 70-80.

Zeren, B., & Özkol, I. (2016). A novel column generation strategy for large scale airline crew pairing problems. *Expert Systems with Applications*, *55*, 133-144.

Zhang, Q., & Li, H. (2007). MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, *11*(6), 712-731.