Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Transportation Research Part C 152 (2023) 104188

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A distributionally robust optimization approach for airline integrated recovery under in-flight pandemic transmission risks

Yifan Xu, Sebastian Wandelt, Xiaoqian Sun *

School of Electronic and Information Engineering, Beihang University, 100191, Beijing, China
National Engineering Laboratory of Multi-Modal Transportation Big Data, 100191, Beijing, China

ARTICLE INFO

Keywords:
Integrated airline recovery
Pandemic
Distributionally robust optimization

ABSTRACT

The COVID-19 pandemic has hit the airline industry hard, leading to heterogeneous epidemiological situations across markets, irregular flight bans, and increasing operational hurdles. Such a melange of irregularities has presented significant challenges to the airline industry, which typically relies on long-term planning. Given the growing risk of disruptions during epidemic and pandemic outbreaks, the role of airline recovery is becoming increasingly crucial for the aviation industry. This study proposes a novel model for airline integrated recovery problem under the risk of in-flight epidemic transmission risks. This model recovers the schedules of aircraft, crew, and passengers to eliminate possible epidemic dissemination while reducing airline operating costs. To account for the high uncertainty with respect to in-flight transmission rates and to prevent overfitting of the empirical distribution, a Wasserstein distance-based ambiguity set is utilized to formulate a distributionally robust optimization model. Aimed at tackling computation difficulties, a branch-and-cut solution method and a large neighborhood search heuristic are proposed in this study based on an epidemic propagation network. The computation results for real-world flight schedules and a probabilistic infection model suggest that the proposed model is capable of reducing the expected number of infected crew members and passengers by 45% with less than 4% increase in flight cancellation/delay rates. Furthermore, practical insights into the selection of critical parameters as well as their relationship with other common disruptions are provided. The integrated model is expected to enhance airline disruption management against major public health events while minimizing economic loss.

1. Introduction

As a result of the rapid growth of air transportation, billions of worldwide passenger flow and close personal contact gradually raise important health concerns regarding epidemic transmission during air travel. Through extensively documented cases, air mobility is well known to serve as a conduit for the rapid dissemination of emerging pandemics (Hertzberg et al., 2018). Representative diseases like influenza, SARS, and MERS (Namilae et al., 2017; Ding et al., 2021) are reported to infect air passengers via contact, airborne, common vehicle, and vector-borne routes (Mangili and Gendreau, 2005). In particular, the recent SARS-CoV—2 coronavirus which causes COVID-19 has spread at planet-scale and unprecedentedly hit the airline industry (Suau-Sanchez et al., 2020; Birolini et al., 2021; Sun et al., 2021b) with frequent reports of in-flight transmission (Freedman and Wilder-Smith, 2020; Government of Canada, 2021).

* Corresponding author at: School of Electronic and Information Engineering, Beihang University, 100191, Beijing, China.
E-mail addresses: yifan_buaa@buaa.edu.cn (Y. Xu), wandelt@informatik.hu-berlin.de (S. Wandelt), sunxq@buaa.edu.cn (X. Sun).

https://doi.org/10.1016/j.trc.2023.104188
Received 10 May 2022; Received in revised form 27 April 2023; Accepted 24 May 2023
Available online 7 June 2023

The major challenges resulted from these highly influential diseases are the ramifications due to sudden flight bans and the downgrade in airline service levels. Flight bans have been frequently used for the purpose of interrupting or slowing down the process of contagion spreading through the air transportation system. In 2014, the outbreak of Ebola in West Africa prompted flight suspensions and border closures from a number of governments and airlines. African countries were predicted to experience over 70% reduction in traffic flow (Poletto et al., 2014). On the other hand, a crew shortage due to a surge in infection rates is inclined to cause flight delays and cancellations. For instance, on January 2nd, 2022, airlines in the United States canceled over 4000 flights due to adverse weather and a sharp increase in Omicron-related cases.[1]

While the regular operation of airlines has already suffered from unforeseen common disturbances like reduced runway capacity, aircraft mechanic failures, and convective weather, the pandemic dissemination-related disruptions further exacerbate the challenges faced by airlines to maintain operational efficiency. Despite the interest of many researchers in the ordinary airline recovery problem against common disruption scenarios, few studies additionally consider the corresponding countermeasures for disease spreading. As specified by European Commission (2020), cancellations due to concerns for crew health and flight restrictions are considered extraordinary circumstances. Airlines consequently owe the chance to improve and recover flight schedules to protect their passengers and crews against diseases.

Pandemics can lead to various heterogeneous influences on aviation, such as demand fluctuation and increased aircraft turnaround time for disinfection (Sun et al., 2021a, 2022b). In this study, we focus particularly on in-flight pandemic dissemination among crews (pilots) and air passengers. The targeted airline integrated recovery problem under in-flight pandemic transmission risk aims to reactively reduce the airline operation cost as well as the overall infection level of crews and passengers within a one-day planning horizon under pandemic disruptions. These disruptions are primarily driven by initial crew and passenger infection when one or more crew members have tested positive and the varying regional infection levels for several passenger originated cities. Although the epidemic transmission risk can be approximated by the basic reproduction number R_0, there exists a large extent of uncertainty for different environments and individuals (Namilae et al., 2017; Lau et al., 2020). Based on in-flight transmission risks data, we model the infection uncertainty through a distributionally robust approach with Wasserstein distance-based ambiguity set. A branch-and-cut and large neighborhood search methods are developed to iteratively solve the optimization model with efficiency. The proposed model and algorithms are validated using a real-world case study under COVID-19-related disruptions, which can also be extended to other pandemic cases.

From a methodological standpoint, this paper provides a comprehensive and data-driven approach to address the scheduling problem involving multiple airline resources with uncertainty in in-flight pandemic transmission. Operating recovery decisions with respect to aircraft routes, crew schedules, and passenger assignment are jointly produced following specific working rules and eligible constraints in the master problem; epidemic propagation routes and risks are identified sequentially in a subproblem based on an epidemic propagation network and empirical infection data. Hence, by jointly optimizing two components dynamically, onboard passengers and aircrews are protected against pathogenic interactions with risky individuals. Integration of multiple resources further leads to global optimal schedules with balanced operational recovery costs and expected infection costs.

From a practical standpoint, the proposed model and algorithms provide decision support for airline recovery optimization under disruptive health security incidences. Based on a collaborative environment between different departments at the airline operations control center, which enables timely information sharing about operating resources and disease dissemination, the complete schedules involving aircraft, crews, and passengers can be modified properly. The proposed approach provides a reliable way to secure the health of aircrews and passengers, accordingly alleviating potential crew shortages due to illness and boosting passenger willingness to travel. Based on a real-world case study, the expected infection number of the adopted approach is found to be 45% lower than that of the practical risky crew isolation strategy, with slight increases in flight cancellation/delay rates (no more than 4%).

The remainder of this study is organized as follows. Section 2 reviews the literature on airline disruption management and the operational responses of airlines during the pandemic and summarizes the contributions of this study. Section 3 presents the formulation of the mathematical model with two solution techniques and an illustrative example. Section 4 performs an extensive set of experiments on a real-world case study, highlighting the benefits of the proposed solution algorithms and integrated recovery model against benchmark contrasting models. Finally, Section 5 concludes this study and provides an outlook for future work directions.

2. Literature review

With the prevalence of several influential diseases, a growing number of papers have been published on the passenger-centric flight experience in the presence of the pandemic (Sun et al., 2021c, 2022a). Besides, there exist extensive studies on airline recovery problem that aim to handle perturbations from both airline resources and the external environment. Research relevant to our work that focuses on improved recovery decisions for in-flight pandemic transmission is collected and briefly reviewed in this section to highlight our main contributions.

[1] https://www.reuters.com/markets/commodities/airlines-grapple-with-omicron-related-disruptions-start-off-2022-2022-01-02/.

Table 1
Overview of research on airline disruption management.

Reference	A/C	Crew	Pax	Delay	Model	Objective	Algorithm
Thengvall et al. (2003)	✓			Flight copies	Arc	Max revenue, -FD, -DS	Lagrangian relaxation
Abdelghany et al. (2004)		✓		Decision variable	Arc	Min crew cost, crew delay	Rolling horizon
Bratu and Barnhart (2006)	✓	✓	✓	Flight copies	Arc	Min PD, PC, RC	B&B
Nissen and Haase (2006)		✓			Flight	Min DS	B&P
Eggenberg et al. (2010)	✓			Flight copies	Path	Min FD, FC	CG
Petersen et al. (2012)	✓	✓	✓	Flight copies	Path	Min FD, FC, DS, CDH, CD	CG, Benders decomposition
Sinclair et al. (2014)	✓		✓	Fixed interval		Min FD, FC, PD, DS	LNS
Maher (2015)	✓	✓	✓	Flight copies	Path	Min FD, CDH, RC, passenger reallocation, DS	CRG
Zhang et al. (2015)	✓	✓		Flight copies	Arc	Min FD, FC, CD	Two stage algorithm
Maher (2016)	✓	✓		Flight copies	Path	Min FD, FC, CD, RC, CDH	CRG
Arıkan et al. (2017)	✓	✓	✓	Decision variable	Arc	Min PD, FC, CDH, FD, AF	QCP solver
Marla et al. (2017)	✓		✓	Flight copies	Arc	Min PD, DS, AF	B&B
Liang et al. (2018)	✓			Adaptive	Path	Min FD, FC, DS	CG
Lee et al. (2020)	✓			Flight copies	Arc	Min FD, FC, DS, AF	Look ahead approximation, sample average approximation
Vink et al. (2020)	✓			Flight copies	Arc	Min FD, FC, DS, operation costs	Iterative selection method
Huang et al. (2021)	✓			Flight copies	Path	Min FD, FC, DS	Primal–dual method
Hu et al. (2021)	✓		✓		Arc	Min FD, FC, PD, refunding, willingness failure	Variable neighborhood search
This study	✓	✓	✓	Decision variable	Arc	Min FD, FC, crew infection	Branch-and-cut, LNS

AF: additional fuel, **CD**:crew duty, **CDH**: crew deadhead, **DS**: deviations from schedule;
FD: flight delay, **FC**: flight cancellation, **PC**: passenger cancellation, **PD**: passenger delay;
RC: reserve crew, **B&B**: branch-and-bound, **LNS**: large neighborhood search, **B&P**: branch-and-price;
CG: column generation, **CRG**: column-and-row generation, **QCP**: quadratic conic programming.

2.1. Airline recovery problems

The high complexity drives the complete airline recovery problem to be traditionally separated into several sequential optimization subproblems: aircraft recovery (Thengvall et al., 2003), crew recovery (Lettovskỳ et al., 2000), and passenger recovery (usually studied with other subproblems). However, recent two decades have witnessed increasing interest in the integrated airline recovery problem (Zhang et al., 2015) encompassing two or more subproblems as a significant amount of additional tactical cost can be saved (Evler et al., 2021). The related literature on airline disruption management is summarized in Table 1, where column *A/C* and *Pax* denote aircraft and passengers, respectively. The column *Model* demonstrates the modeling choices where the arc and flight-based models feature a more compact formulation when compared with the path-based model. Finally, in column *Objective*, we illustrate the recovery objectives of the research work with penalties on aircraft swaps, maintenance swaps, and other similar approaches that are classified as deviations from the original schedules.

To efficiently allocate airline resources, **recovery subproblems** related to aircraft and crews are widely studied in the literature and the aircraft recovery problem is particularly well studied due to the successive dependency of crew rescheduling and passenger resettlement (Su et al., 2021). While flight delay and cancellation are common recovery decisions in the literature, the way in which flight delay is modeled (e.g. flight copies, decision variables) determines the performance of the resulting model formulation. To optimize aircraft recovery against a hub closure, Thengvall et al. (2003) use a multi-commodity flow model to maximize the revenue less the recovery cost. A Lagrangian relaxation-based bundle algorithm is presented to generate near-optimal heuristic solutions. Abdelghany et al. (2004) present a decision support tool to optimize crew recovery for hub-and-spoke airlines. In particular, crew problems are proactively reserved before disruption with the purpose of flexibility. Different from the commonly applied flight copy approach using fixed intervals, Liang et al. (2018) and Huang et al. (2021) propose two alternative solution methodologies to set partition and compact models, respectively. The aircraft recovery model in Liang et al. (2018) considers airport

capacity and maintenance flexibility. Continuous delay is incorporated through a column generation-based heuristic. On the other hand, in Huang et al. (2021), the aircraft recovery model uses a cost-driven copy generation approach for flight and maintenance re-timing. Based on a copy evaluation, a primal–dual approach is used by iteratively adding high quality flight copies.

Contrary to the sequential subproblems, **integrated recovery problems** aim to achieve solutions considering the coupling relationship between different resources and feature reduced operation costs but with drastically increasing problem complexity. To start with, Bratu and Barnhart (2006) propose a model to balance airline operating costs and passenger delay costs. Recovery decisions including flight re-timing, cancellation, swapping, and reserving crews are incorporated and compliance with crew regulations and aircraft maintenance requirements is ensured. Petersen et al. (2012) propose an optimization model to derive passenger-friendly integrated solutions considering aircraft, crew, and passengers. To yield solutions with reasonable computational efforts, Benders decomposition and column generation algorithms are applied to the restricted model. Whilst traditional branch-and-price algorithm does not provide satisfactory computational performance for the integrated airline recovery problem, Maher (2015) presents a MIP model considering passenger reallocation cost with knapsack variables, a column-and-row generation algorithm is developed to avoid enumerating decision variables and flight copy-related constraints. Emphasizing the importance of cruise speed adjustment, Arıkan et al. (2017) propose a quadratic optimization model with explicit modeling of aircraft, crew, and individual passenger recovery resolutions under cruise speed control. Two pre-processing techniques are presented to eliminate unnecessary variables and constraints related to the flight connection network. Lastly, Hu et al. (2021) present an integrated MIP model that optimizes aircraft recovery cost and passenger cost involving willingness using prospect theory. The multi-objective model is solved with a heuristic variable neighborhood search algorithm, and its efficiency is validated through a comparison with the local search method.

In brief, handling aircraft, crew, and passengers in concrete generally results in global optimal solutions as compared to the sequential process where solution fixing is unavoidable at each stage. The arduous task of solving integrated recovery problems, in turn, consistently requires new scalable methodologies to repair airline schedules in a timely manner.

2.2. Airline operational responses to epidemic spreading

To efficiently handle the epidemic transmission risk of air travel, there is an increasing research interest in enhancing the relevant airline procedures. In particular, operations involving flight turnarounds are mainly studied given the fact that infection probability for passengers seated beyond two rows from an infectious individual is relatively low (Hertzberg and Weiss, 2016). Specifically, to maintain the social distance among air passengers, two deterministic models are proposed in Salari et al. (2020) based on two distance metrics: keeping the passengers seated far enough away from each other and keeping a safe distance from the aisle. Similarly, Schultz and Soolaki (2021) optimize seat allocation and sequence of boarding based on passenger groups. A mathematical model is developed accordingly to provide solutions to minimize the sum of shedding rates, which directly reduces the infection probability caused by an infected passenger. To account for the additional time required for cabin disinfection and passenger boarding, Schultz et al. (2020) develop a resource-constrained cabin cleaning scheduling model with disinfection operation to minimize the makespan of the complete process. This model is integrated into a turnaround reference model with stochastic boarding simulation to study the impact on total turnaround time. In the work of Shafipour-Omrani et al. (2021), a bi-objective crew pairing model is proposed and the risk of pandemic infection is implicitly reduced by decreasing the total elapsed time. This may lead to a shorter sit time and fewer flights per duty, resulting in less time spent on catering. In reality, a few airlines have already introduced indicators to limit the number of crew members that an individual comes in contact with Garrow et al. (2022).

2.3. Contributions

The current optimization studies on airline recovery problems are mostly motivated by disruptions of airline resources or the external environment, while in-flight epidemic spreading related risks have not been addressed. On the other hand, existing studies on operational responses to pandemics mainly focus on passenger seat allocation on a single flight or indirectly optimizing the crew schedule. However, limited effort has been devoted to taking pandemic transmission risks into account for operational recovery reactions where the propagation effects of contagious diseases are partially overlooked given frequent passenger transfers and crew swapping opportunities. Accordingly, there are several important areas where this study makes an original contribution to the literature:

1. We combine an integrated airline recovery model with an epidemiological simulation component, which is able to derive better and more reliable schedules under the situation of unevenly distributed pandemic states across served markets aside from other commonly discussed disruption scenarios like airport closures and delayed ready time. In contrast to existing recovery models that integrate crew members into crew teams to reduce complexity, our model considers the personalized crew schedule and passenger itineraries with epidemic propagation. The novel model can cut off epidemic transmission paths at an early stage and protect crew members/passengers effectively with reduced social and operating costs.

2. We model the individual infection probability as a random variable that is well known to vary among different variants, genders, and prevention measures (e.g. mask-wearing condition, seat locations). Specifically, it is assumed to follow a probability distribution that is only partially known within a Wasserstein distance-based ambiguity set. This approach facilitates data-driven modeling of infection behaviors, reaching a trade-off between widely applied sample average approximation and robust optimization (uncertainty set).

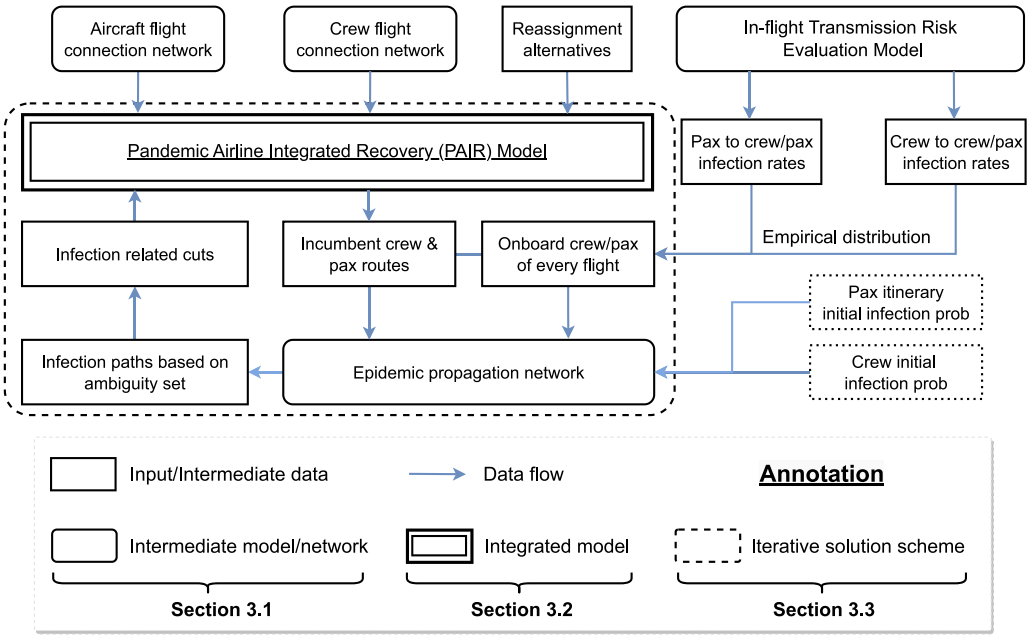


Fig. 1. The modeling architecture of the airline integrated recovery problem under in-flight pandemic transmission risk.

3. We propose a branch-and-cut and large neighborhood search-based solution framework to efficiently solve the pandemic airline integrated recovery model. These algorithms jointly help to avoid enumerating high-dimensional crew/passenger coupling constraints and their resulting computation intractability. Extensive experiments are performed using real-world datasets to demonstrate the performance of our model and the solution methodology against deterministic and robust optimization models. Sensitivity analyses are also carried out to identify the appropriate selection of the ambiguity set and the infection-related cost. Finally, operational distinctions between infection and other types of disruption are investigated.

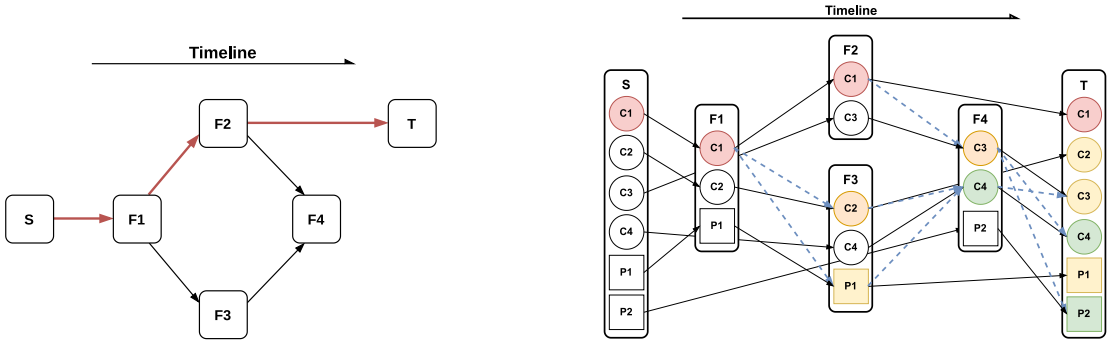
3. Mathematical formulation and methodologies

As illustrated in Fig. 1, the pandemic airline integrated recovery(PAIR) model serves as the primary element of the overall modeling architecture. Flight connection networks for different individual aircraft and crew members along with passenger reassignment alternatives are first created. Then the model is solved with an iterative solution approach, in which an epidemic propagation subproblem is optimized to obtain the cuts that represent the maximum expected infection cost of the incumbent crew and passenger schedule based on the epidemic propagation network. The edge weights of the network are derived from an in-flight transmission risk evaluation model based on specific attributes like cabin configuration, pandemic statistics, and mask usage. Hence, epidemic infection paths can be identified with uncertain transmission rates, which are modeled through an ambiguity set.

The main novelty of the PAIR model is the embedding of an epidemic propagation subproblem to minimize the maximum expected infection cost of the current crew and passenger assignment. Based on the assignment, entities on the same flight, which refer to crew members and passenger itineraries, are first collected to identify the possible infection opportunities associated with social contact. Secondary infected entities may induce further dissemination by contacting others during their subsequent flight trips. Considering the differences in alternative seat positions and population sizes of entity types (one for crew members and actual itinerary demand otherwise), four types of transmission risks are generated according to an in-flight infection risk evaluation model to serve as empirical distributions. Therefore, the subproblem enables microscopic epidemic propagation within one entity type: crew-to-crew, passenger-to-passenger, or different entity types: crew-to-passenger, passenger-to-crew (see Appendix A for more details of transmission risk evaluation).

Before proceeding further, it is noteworthy that individual crew members and passenger itineraries are adopted to model the rescheduling decisions and pandemic propagation at different granularity levels due to the problem's complexity. For instance, a realistic scheduling instance with 81 flights contains 69 pilots along with 12,764 passengers. Thus, aggregating passengers into itineraries helps to derive a more scalable model formulation but partially overlooks the disease dissemination effects among passengers on the same itinerary, which can be alleviated given appropriate scattered seat distributions.

In this section, the underlying flight connection network and epidemic propagation network are first presented in Section 3.1. The mathematical model is then introduced in Section 3.2 with the necessary reformulation based on the Wasserstein distance ambiguity set. The exact branch-and-cut and a large neighborhood search heuristic are presented in Section 3.3. The benefits of the PAIR model are exemplified in Section 3.4 in contrast to an ordinary recovery model without infection-related optimization.



(a) Flight connection network for crew **C1** (b) Epidemic propagation network

Fig. 2. The underlying networks of PAIR model. Different circle/rectangle colors in subfigure (b) represent the order of infection.

3.1. Network structures

In order to model aircraft and crew routing decisions, the flight connection network *G(V, E)* is constructed for each individual aircraft and crew member. To be more specific, every vertex $v \in V$ denotes a flight while every edge $e \in E$ represents a possible valid flight connection if the arrival airport of the first flight is the same as the departure airport of the second flight and the arrival time of the first flight minus a given time threshold φ is less than the departure time of the second flight to allow for flexibility in flight delay decisions. Two dummy source and sink vertices are added additionally to represent the origin and destination airports of a specific aircraft or crew member and are connected to flights departing from or arriving at these airports. Fig. 2(a) demonstrates a flight connection network for crew *C1* encompassing four flight vertices (from *F1* to *F4*) with the source *S* and sink *T* vertices representing his start/end airports. Red lines are used to illustrate one possible schedule of *C1* where *F1* and *F2* are consecutively executed in increasing time order as indicated by the timeline.

Given the complete schedule which includes crew routing and passenger itineraries, an epidemic propagation network can be constructed where each vertex indicates a combination of a crew/passenger itinerary and a flight with two dummy vertices introduced for each crew and itinerary to represent the initial and final states. Non-stop and one-stop passenger itineraries (i.e. a flight sequence serving a specific OD market) are considered in this context because itineraries with more stops only make up 2.5% of one-way trips (Barnhart et al., 2014). Two types of edges are used accordingly with varying weights for infection probabilities. The first type of edge is constructed between the same **entity** (a crew or passenger itinerary) with an infection transmission probability of one. While the second type of edges is constructed among different entities to represent the potential disease transmission with uncertain infection probabilities due to their previous staying on the same flight. For instance, the infection probability from entity (crew/passenger itinerary) e_1 to entity e_2 can be computed as $\tau_{e_1}^{f}\beta_{e_1e_2}$ where the constant parameter $\tau_{e_1}^{f} = N_{e_1}\frac{FT_f}{60}^{k_1}$ represent the impact of infector number of entity i and the elapsed time (min) of flight f on infection with a scaling factor of k_1. $\beta_{e_1e_2}$ is the probability that individual e_2 is infected by one infector of entity e_1 during a 60-min flight considering their alternative seat locations. Since such value depends not only on social patterns of contact but also on the pathology of the infectious organism, a probabilistic infection evaluation model for in-flight epidemic transmission is adopted to accurately generate the data samples as illustrated in Appendix A.

As depicted in Fig. 2(b), four crew members (*C1–C4*) and two passenger itineraries (*P1–P2*) are assigned to flights and two types of edges are indicated by solid black lines and dashed blue lines, respectively. Assuming crew *C1* has previously contacted one confirmed patient and is risky to infect other crew members as close contacts. In this case, crew *C2* is likely to be infected through the blue arc and spread the virus afterwards to infect crew *C4* on flight *F3* given the fact that a person can be contagious even during the incubation period. Similarly, crew *C3* may also be infected by *C1* during their contact on flight *F2*. The one-stop itinerary also facilitates the propagation of the epidemic from flight *F1* to flight *F3*, indicating possible virus spreading opportunities between crew members and passengers.

Under normal operation, the same crew group consisting of multiple crew members is consistently used to execute a complete duty/pairing which will not result in many crew intersections as shown in the figure. However, this may not be efficient if there is a crew shortage due to delayed ready time or illness. In this case, more flexible use of different crews is required to reduce cancellations and delays. Epidemic infection routes $r \in R_{C3}$ for *C3*, for instance, can be found from source vertex $\langle S, C1\rangle$ to sink vertex $\langle T, C3\rangle$ as follows.

$$\{\langle S, C1\rangle, \langle F1, C1\rangle, \langle F2, C1\rangle, \langle F4, C3\rangle, \langle T, C3\rangle\}$$
$$\{\langle S, C1\rangle, \langle F1, C1\rangle, \langle F3, C2\rangle, \langle F4, C4\rangle, \langle T, C3\rangle\}$$

Besides, notation $(f, e_1, e_2) \in r$ denotes the uniquely visited flights and entities in sequence. In this case, $r_1 = \{(F2, C1, C3)\}$ and $r_2 = \{(F1, C1, C2), (F3, C2, C4), (F4, C4, C3)\}$.

Table 2
Sets, parameters and decision variables for the PAIR model.

Sets			
F	Set of flights, $f \in F$.	T	Set of tails (aircraft), $t \in T$.
C	Set of crew members, $c \in C$.	I	Set of passenger itineraries, $i \in I$.
S^t/T^t	Source/sink vertex in the flight connection network of aircraft(crew) t.	F_i/FP_i	Set of flights/flight pairs in itinerary i.
E	Set of edges in flight connection network, $(f,g) \in E$.	I_i^R	Set of alternative redirecting itineraries of itinerary i.
I_f	Set of itineraries traversing flight f.	$R_e(\mathbf{x},\mathbf{s})$	Set of infection routes for entity e given solution $\mathbf{x},\mathbf{s}$, $r = \{r_0, r_1, \ldots\} \in R_e(\mathbf{x},\mathbf{s})$.
Parameters			
Req_f	Number of crews required to operate flight f.	FT_f	Block time for flight f.
CT_{fg}	Minimum connection time between flight f and g.	DT_f	Time of departure for flight f.
AT_f	Time of arrival for flight f.	MD	Maximum delay time.
tc_f^1	Cancellation cost of flight f.	tc_f^2	Delay cost of flight f.
tc_f^3	Cost for each crew to use a deadhead flight f.	tc_i^4	Unassigned cost for passenger in itinerary i.
tc_e^5	Infection cost of entity e.	MFT_c	Maximum credit flight time per duty for crew c.
MLN_c	Maximum number of landings per duty for crew c.	D_i	Number of passengers on itinerary i
S	Available seats of aircraft.	PCT_{fg}	Minimum passenger connection time between flight f and g.
$M/M^1/M^2$	Big number.		
Decision variables			
$dt_f \geq DT_f$	Actual departure time of flight f.	$at_f \geq AT_f$	Actual arrival time of flight f.
$x_{fg}^c/x_{fg}^t \in \{0,1\}$	1 if crew c/tail t use edges (f, g), 0 otherwise.	$z_f \in \{0,1\}$	1 if flight f is cancelled, 0 otherwise.
$d_f \geq 0$	Delay time of flight f.	$u_f \geq 0$	Number of crews take deadhead flight f.
$b_i^j \geq 0$	Number of passengers redirected from itinerary i to j.	$s_i \in \{0,1\}$	1 if the itinerary i is not disrupted, 0 otherwise.
$w_{fg} \in \{0,1\}$	1 if enough time between flight f and g is reserved for passenger transfer.		

Based on these terminologies, the overall infection risk η_e of entity e is computed with Eq. (1) given the initial infection probability of the first risky entity of route r as $\beta_{r_0}^0$.

$$\eta_e = \max_{r \in R_e} \beta_{r_0}^0 \prod_{(f,e_1,e_2) \in r} \tau_{e_1}^f \beta_{e_1,e_2} \tag{1}$$

In this equation, R_e is the set of routes that end at $\langle T, e \rangle$ and each route r is composed of a range of visited crew members/passenger itineraries with their executing flights from the source vertex (initial risky entities) to the sink vertex.

In terms of passenger-related infections, due to the contact between secondary risky crew *C1* and itinerary *P1* on flight *F1*, the propagated infection probability is computed with the multiplication of the initial infection rate *C1*, the crew-to-passenger transmission rate $\beta_{C1,P1}$ and the constant parameter τ_{C1}^f where $N_{C1} = 1$. Afterwards, the close contact of itinerary *P1* with crew *C4* enables further epidemic propagation depending on passenger-to-crew transmission rate $\beta_{P1,C4}$ and τ_{P1}^f with population size D_{P1}. Since *C4* may be affected by both *C2* and *P1*, the maximum product value is selected according to Eq. (1).

With the unified concept of entities and infection edges, different epidemic transmission categories between crews and passenger itineraries are generalized and represented in the epidemic propagation network, which therefore allows microscopic modeling of in-flight disease dissemination.

3.2. Mathematical formulation

On the basis of the compact formulation proposed by Arıkan et al. (2017), the integrated recovery model includes commonly applied decisions of flight delays, cancellation, aircraft swap, and crew re-routing. Additionally, crew duty constraints, passenger reassignment, and interrelated infection propagation factors are newly incorporated using the distributionally robust optimization approach to model uncertainty in the transmission probability. In this section, the complete nonlinear model is first presented then it is reformulated into a mixed-integer model based on duality theory.

3.2.1. The pandemic airline integrated recovery model

Given the relevant notation regarding sets, parameters, and decision variables presented in Table 2, the complete pandemic airline integrated recovery (PAIR) model is illustrated in (2)–(21).

$$\begin{aligned}(PAIR)\ \min \sum_{f \in F}(tc_f^1 z_f + tc_f^2 d_f + tc_f^3 u_f) + \sum_{i \in I} tc_i^4 (D_i - \sum_{j \in I_i^R} b_i^j) + \\ \sum_{e \in C \cup I} \left(tc_e^5 \max_{r \in R_e(\mathbf{x},\mathbf{s})} \beta_{r_0}^0 \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}[\prod_{(f,e_1,e_2) \in r} \tau_{e_1}^f \beta_{e_1,e_2}] \right)\end{aligned} \tag{2}$$

$$\sum_{f:(f,g)\in E} x^t_{fg} - \sum_{f:(g,f)\in E} x^t_{gf} = \begin{cases} -1 & g = S^t \\ +1 & g = T^t \\ 0 & \text{o/w} \end{cases} \qquad \forall t \in T \cup C, g \in V \tag{3}$$

$$\sum_{t\in T}\sum_{g:(f,g)\in E} x^t_{fg} = 1 - z_f \qquad \forall f \in F \tag{4}$$

$$\sum_{c\in C}\sum_{g:(f,g)\in E} x^c_{fg} - u_f = (1 - z_f)Req_f \qquad \forall f \in F \tag{5}$$

$$\sum_{c\in C}\sum_{g:(f,g)\in E} x^c_{fg} \le M_f(1 - z_f) \qquad \forall f \in F \tag{6}$$

$$at_f = dt_f + \sum_{t\in T}\sum_{g:(f,g)\in E} x^t_{fg} FT_f \qquad \forall f \in F \tag{7}$$

$$dt_g \ge at_f + CT_{fg}x^t_{fg} - (AT_f + MD)(1 - x^t_{fg}) \qquad \forall t \in T \cup C, (f,g) \in E \tag{8}$$

$$d_f \ge at_f - AT_f - M_f z_f \qquad \forall f \in F \tag{9}$$

$$\sum_{f\in F}\sum_{f:(f,g)\in E} FT_f x^c_{fg} \le MFT_c \qquad \forall c \in C \tag{10}$$

$$\sum_{f\in F}\sum_{f:(f,g)\in E} x^c_{fg} \le MLN_c \qquad \forall c \in C \tag{11}$$

$$\sum_{c\in C}\sum_{g:(f,g)\in E} x^c_{fg} \ge u_f \qquad \forall f \in F \tag{12}$$

$$\sum_{i\in I}\sum_{j\in I_f, j\in I^R_i} b^j_i \le S(1 - z_f) \qquad \forall f \in F \tag{13}$$

$$\sum_{j\in I^R_i} b^j_i \le D_i \qquad \forall i \in I \tag{14}$$

$$\sum_{i\in I, i\in I^R_j} b^j_i \le Ss_j \qquad \forall j \in I \tag{15}$$

$$b^i_i \ge D_i s_i \qquad \forall i \in I \tag{16}$$

$$dt_g - at_f \ge PCT_{fg} - M^1_{fg}(1 - w_{fg}) \qquad \forall i \in I, (f,g) \in FP_i \tag{17}$$

$$dt_g - at_f \le PCT_{fg} + M^2_{fg}w_{fg} \qquad \forall i \in I, (f,g) \in FP_i \tag{18}$$

$$s_i \le 1 - z_f \qquad \forall i \in I, f \in F_i \tag{19}$$

$$s_i \le w_{fg} \qquad \forall i \in I, (f,g) \in FP_i \tag{20}$$

$$s_i \ge \sum_{(f,g)\in FP_i} (w_{fg} - 1) - \sum_{f\in F_i} z_f + 1 \qquad \forall i \in I \tag{21}$$

The objective function (2) minimizes the weighted costs of flight cancellation, delays, crew deadhead, unassigned passengers, and entity infection. The third term of supreme expected infection cost in ambiguity set $\mathcal{F}$ relies on a subproblem that derives the infection paths from the epidemic propagation network given a complete crew schedule $\mathbf{x}$ and valid passenger routes $\mathbf{s}$ which will be further explained. Flow balance Constraints (3) ensure the feasible routing of crews and aircraft. Flight cover Constraints (4) and (5) ensure every flight is executed if aircraft and enough crews are assigned and crew deadhead is incorporated by allowing more than Req_f crews to be carried. Otherwise, Constraints (6) specify that no crew can be transported on flight f. Constraints (7) ensure that sufficient flight time is reserved between the departure and arrival time of each flight if not canceled. Flight connection time is captured in Constraints (8) and flight delays are calculated in Constraints (9). The maximum credit flying time of each crew member is ensured with Constraints (10) and the maximum number of landings allowed per duty is ensured in Constraints (11). Constraints(12) ensure that if any crew is carried with a deadhead flight then the correlated flight edge must be used.

Passenger reassignment decisions are modeled with Constraints (13)–(20). Based on a pre-selected set of redirect alternatives I^R_i for passenger itinerary i ($i \in I^R_i$), the left-hand side sums of Constraints (13) derive the actual number of passengers carried on every flight f which should not exceed the aircraft capacity if the flight is not canceled. Besides, the redirected passenger number from itinerary i is restricted by Constraints (14) to the total passenger demand D_i. The feasibility of any itinerary s_i relies on the execution status of its component flights z_f as well as the connection time between each successive flight pair w_{fg}. Constraints (15) ensure no passenger is allocated to any invalid itinerary and all passengers should be kept on their original itineraries if no cancellation or miss-connection happens as specified by Constraints (16). Big-M Constraints (17) and (18) are applied to decide whether sufficient connection time is reserved between flight f and g. When $w_{fg} = 1$, Constraints (17) ensure the time interval must be larger than the minimum passenger connection time whilst Constraints (18) become redundant. On the other hand, Constraints (18) ensure the flight pair (f, g) cannot be used for passenger connections. Finally, the upper and lower bound of s_i are ensured by Constraints (19)–(21).

3.2.2. Mixed-integer linear reformulation

The complexity of the mathematical model lies in the infection-related decisions that are embedded in the definition of $R_e(\mathbf{x})$. Specifically, given an integer feasible solution $\mathbf{x}$ and $\mathbf{s}$, the epidemic propagation routes r with values of $\sup_{\mathbb{P}\in F}\mathbb{E}[\prod_{(f,e_1,e_2)\in r}\tau_{e_1}^f\beta_{e_1,e_2}]$ can be derived by solving the cut generation subproblem with the acyclic graph search algorithm. For each route r, the infection cost in the objective function (2) can be rewritten as $\exp\{\sum_{(f,e_1,e_2)\in r}\ln\tau_{e_1}^f+\ln\beta_{e_1,e_2}\}$. Let $\alpha_{e_1,e_2}=\ln\beta_{e_1,e_2}$, due to the aforementioned uncertainty in infection probabilities, neglecting them may lead to either an over-optimistic or conservative attitude that consequently results in inefficient and unreliable crew recovery resolutions. Therefore, in this study, we adopt a distributionally robust optimization approach that optimizes the schedule under the supremum expectation of $\exp\{\sum_{(f,e_1,e_2)\in r}\tilde{\alpha}_{e_1,e_2}\}\cdot\exp\{\sum_{(f,e_1,e_2)\in r}\ln\tau_{e_1}^f\}$ to reformulate the model and approximate that of $\prod_{(f,e_1,e_2)\in r}\tau_{e_1}^f\beta_{e_1,e_2}$ for computation efficiency, given an ambiguity set that prevents overfitting solutions to the empirical distribution.

In this context, we investigate a Wasserstein distance-based ambiguity set that is computationally efficient and does not assume an absolute continuous distribution (Zhang et al., 2021). While the exact distribution of $\tilde{\boldsymbol{\alpha}}$ cannot be known, empirical log virus infection probabilities $\hat{\boldsymbol{\alpha}}_\omega$ are accessible from either historical data or certain prediction models. We denote $\omega\in\Omega$ as the empirical values in scenario ω. Then, as shown in Eq. (22), the empirical distribution $\mathbb{P}^\dagger$ is derived from the uniform distribution over the historical data (N samples).

$$\mathbb{P}^\dagger[\tilde{\boldsymbol{\alpha}}=\hat{\boldsymbol{\alpha}}_\omega]=\frac{1}{N} \tag{22}$$

Subsequently, we define the Wasserstein ambiguity set using parameter θ as follows

$$\mathcal{F}(\theta)=\left\{\mathbb{P}\in\mathcal{P}(\mathcal{W})|\tilde{\boldsymbol{\alpha}}\sim\mathbb{P},\tilde{\boldsymbol{\alpha}}^\dagger\sim\mathbb{P}^\dagger,d_{\mathcal{W}}(\mathbb{P},\mathbb{P}^\dagger)\le\theta\right\} \tag{23}$$

$$d_{\mathcal{W}}(\mathbb{P},\mathbb{P}^\dagger)=\inf_{\bar{\mathbb{P}}}\mathbb{E}_{\bar{\mathbb{P}}}\left[\|\tilde{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}^\dagger\|_p\right] \tag{24}$$

In the ambiguity set (23), $\mathcal{P}(W)$ denotes the set of all probability distributions over support W. θ controls the radius of the ball around the empirical distribution $\mathbb{P}^\dagger$ and the distance is defined using the transportation cost function d_W as stated in Eq. (24) which calculates the expected p norm value over the joint distribution. In particular, a joint distribution $\bar{\mathbb{P}}$ is constructed over support set $\mathcal{W}\times\mathcal{W},\mathcal{W}=\{\tilde{\alpha}_{e_1,e_2}\ge\underline{\alpha}_{e_1,e_2}\}$ with its marginal distributions being $\mathbb{P}$ and $\mathbb{P}^\dagger$, respectively. Then let $\sum_{(e_1,e_2)}v_{e_1,e_2}\tilde{\alpha}_{e_1,e_2}=\mathbf{v}^T\tilde{\boldsymbol{\alpha}}$ where $v_{e_1,e_2}=1$ if $(e_1,e_2)\in r$ and 0 otherwise, we can evaluate this term as follows.

$$Z=\sup\mathbb{E}_{\bar{\mathbb{P}}}[\mathbf{v}^T\tilde{\boldsymbol{\alpha}}] \tag{25}$$

$$\text{s.t.}\,\mathbb{E}_{\bar{\mathbb{P}}}\left[\|\tilde{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}^\dagger\|_p\right]\le\theta \tag{26}$$

$$(\tilde{\boldsymbol{\alpha}},\tilde{\boldsymbol{\alpha}}^\dagger)\sim\bar{\mathbb{P}} \tag{27}$$

$$\tilde{\boldsymbol{\alpha}}\sim\mathbb{P} \tag{28}$$

$$\tilde{\boldsymbol{\alpha}}^\dagger\sim\mathbb{P}^\dagger \tag{29}$$

$$\bar{\mathbb{P}}[(\tilde{\boldsymbol{\alpha}},\tilde{\boldsymbol{\alpha}}^\dagger)\in\mathcal{W}\times\mathcal{W}]=1 \tag{30}$$

By utilizing the empirical distribution and conditional probability, this problem can be equivalently transformed as

$$Z=\sup\frac{1}{N}\sum_{\omega\in\Omega}\int_{\mathcal{W}}\mathbf{v}^T\boldsymbol{\alpha}\,\mathrm{d}F_\omega(\boldsymbol{\alpha}) \tag{31}$$

$$\text{s.t.}\,\frac{1}{N}\sum_{\omega\in\Omega}\int_{\mathcal{W}}\|\tilde{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}^\dagger\|\,\mathrm{d}F_\omega(\boldsymbol{\alpha})\le\theta \tag{32}$$

$$\frac{1}{N}\int_{\mathcal{W}}\mathrm{d}F_\omega(\boldsymbol{\alpha})=\frac{1}{N}\qquad\forall\omega\in\Omega \tag{33}$$

$$\mathrm{d}F_\omega(\boldsymbol{\alpha})\ge0\qquad\forall\omega\in\Omega \tag{34}$$

where $\mathrm{d}F_\omega(\boldsymbol{\alpha})$ is the cumulative distribution function of infection probabilities for scenario ω in the empirical distribution. Constraints (33) are adjusted by multiplying $\frac{1}{N}$ to facilitate further deduction. The closed-form solution to this problem is given by the following proposition.

Proposition 1. *The optimal closed-form solution of the model Z for entity infection route r is defined as*

$$Z^*=\theta|r|^{\frac{p-1}{p}}+\frac{1}{N}\sum_{\omega\in\Omega}\mathbf{v}^T\hat{\boldsymbol{\alpha}}_\omega \tag{35}$$

Proof. See Appendix B. □

Thereafter, the original nonlinear PAIR model is able to be reformulated as:

$$(PAIR)\ \min\sum_{f\in F}(tc_f^1z_f+tc_f^2d_f+tc_f^3u_f)+\sum_{i\in I}tc_i^4(D_i-\sum_{j\in I_i^R}b_i^j)+\sum_{e\in C\cup I}tc_e^5\eta_e \tag{36}$$

s.t.(3)-(21)

$$\Big(\sum_{x\in E^x_{r_e}}(x^e_{fg}-1)+\sum_{s\in E^s_{r_e}}(s_i-1)+1\Big)\rho_e(\mathbf{x}',\mathbf{s}',\tilde{\boldsymbol{\alpha}})\le\eta_e \qquad \forall e\in C\cup I,\mathbf{x}'\in \mathrm{X},\mathbf{s}'\in \mathrm{S} \tag{37}$$

Here, $E^x_{r_e}, E^s_{r_e}$ denotes the relevant set of variable x, s for each route r_e, $\rho_e(\mathbf{x}',\mathbf{s}',\tilde{\boldsymbol{\alpha}})=\exp\{Z^*+\sum_{(f,e_1,e_2)\in r_e}\ln\tau^f_{e_1,e_2}\}$ represents the infection probability of entity e where Z^* is the closed-form solution according to the current values of $\mathbf{x}'$ and $\mathbf{s}'$.

3.3. Solution methodology

Although the developed PAIR model is reformulated into a MIP model, the exponential number of all feasible solutions hinders efficient solving. Constraints (37) serve as no-good cuts to restrict the lower bound for the current tentative assignment solution of the model. To see this, when the same partial route is selected for entity e, the left-hand side is equal to $\rho_e(\mathbf{x}',\mathbf{s}',\tilde{\boldsymbol{\alpha}})$. Otherwise, the left-hand side value is nonpositive, and the corresponding constraint is redundant.

Hence, this model can be solved with the branch-and-cut algorithm to avoid enumeration of any feasible routing solution X using the callback function of general-purpose solvers such as CPLEX. After deriving integer solution $\mathbf{x}$, the infection propagation graph is first constructed in the subproblem as discussed in Section 3.1. Since the graph is acyclic (edge connections follow increasing time order), the acyclic multi-label graph searching algorithm from Xu et al. (2021) is applied to obtain the longest routes from a dummy source node S via $\langle S,e\rangle, \forall e\in C\cup I$ to each crew/itinerary target vertex $\langle T,e\rangle$.

In particular, vertices are iterated in topological order and two index labels are created to track the cumulation of $\frac{1}{N}\sum_{\omega\in\Omega}\mathbf{v}^T\hat{\boldsymbol{\alpha}}_\omega+\mathbf{v}^T\ln\boldsymbol{\tau}$ and the number of different entities encountered $|r|$. The actual cost of each label is computed as $\theta|r|^{\frac{p-1}{p}}+\frac{1}{N}\sum_{\omega\in\Omega}\mathbf{v}^T\hat{\boldsymbol{\alpha}}_\omega+\mathbf{v}^T\ln\boldsymbol{\tau}$. The edge weight for connection between the same entity is set to $\ln 1=0$ and the edge weight from the dummy source node to any vertex of entity e is set to $\ln\beta^0_e$. To accelerate the algorithm, dominance rules are also applied by eliminating labels with strictly smaller costs and $|r|$. In this way, we derive the infection probability of entity e as $\rho=\exp\upsilon_e$ where υ_e is the value of the longest path found with the multi-label algorithm. With this algorithm, multiple routes (including the shortest one) for every entity can be generated at a single iteration which helps to improve the convergence.

While such cut provides realization of crew/passenger infection propagation quantification based on integer solution of variables $\mathbf{x}$ and $\mathbf{s}$, it leads to enumeration with slow convergence (Chu and Xia, 2004). Therefore, we further strengthen Constraints (37) by introducing intermediate binary variables $y^c_f\in\{0,1\}$ to indicate whether vertex f in the connection network is executed by crew c. The linking relationship between $\mathbf{x}$ and $\mathbf{y}$ is given in Constraints (38).

$$y^c_f=\sum_{g:(f,g)\in E}x^c_{fg} \qquad \forall f\in V,\forall c\in C \tag{38}$$

In this way, the generated cuts for entity e can be rewritten in Constraints (39).

$$\Big(\sum_{y\in E^y_{r_e}}(y^e_f-1)+\sum_{s\in E^s_{r_e}}(s_i-1)+1\Big)\rho_e(\mathbf{y}',\mathbf{s}',\tilde{\boldsymbol{\alpha}})\le\eta_e \qquad \forall e\in C\cup I,\mathbf{y}'\in \mathrm{Y},\mathbf{s}'\in \mathrm{S} \tag{39}$$

Take a passenger infection route from crew c_1 to itinerary i_1 (see Fig. 3) with value $\eta_{i_1}=0.5$ for instance, the cut (37) is expressed as $0.5[(x^{c_1}_{S,f_1}-1)+(x^{c_1}_{f_1,f_2}-1)+(x^{c_1}_{f_2,f_3}-1)+(x^{c_2}_{f_3,f_4}-1)+(x^{c_2}_{f_4,f_5}-1)+(s_{i_1}-1)]+0.5\le\eta_{i_1}$. This cut will be invalidated with a small modification as illustrated with the dashed lines. If crew c_1 executes f_1, f_7, f_3 in sequence, propagation will still effectively happens between c_1 and c_2. Similarly, there also exist opportunities for c_2 to use edge between flight f_3 and f_6 to maintain the same η_{i_1}. However, with the improved expression, the cut can be written as $0.5[(y^{c_1}_{f_3}-1)+0.5(y^{c_2}_{f_3}-1)+0.5(y^{c_2}_{f_5}-1)+0.5(s_{i_1}-1)]+0.5\le\eta_{i_1}$. If the crew infection route only consists of one crew member c_1 the cut can be written as $0.5(y^{c_1}_S-1)\le\eta_{c_1}$. Multiple infection sub-paths ought to be added as cuts as well to accelerate the solution process if more than two crew members are involved.

Finally, we present an effective large neighborhood search algorithm to further accelerate the branch-and-cut algorithm especially when there exist an exponential number of feasible routings for large-scale recovery problems.

The solution algorithm consists of destroying and improvement phases that iteratively improve the integer feasible solutions. To be specific, we divide the complete planning horizon and crew members into smaller groups (neighborhoods) according to their departure time and crew base information, respectively. Thus, the proposed algorithm efficiently solves multiple reduced problems to ensure the full set of resources (flights, aircraft, and crews) are still included in the model. The algorithm design idea is widely applied in the literature (Mancini, 2016; Ribeiro et al., 2019).

The main procedure of the developed large neighborhood search algorithm is summarized as follows:

Step 1: Generating initial feasible solutions. Due to the inclusion of flight cancellation and crew deadhead decisions, the developed PAIR model is always feasible by canceling all involved flights. Consequently, we can start with an integer solution generated from the presented branch-and-cut given a maximum running time of $|F|/2$ in seconds as this time limit provides relatively sufficient global information regarding promising crew connections. Afterward, we set the maximum running time to be $|F|/4$ to quickly solve the reduced problems using the commercial solver and record the best-known lower bound value.

Step 2: Fix and improve neighborhoods. Our neighborhood construction process relies on the idea that nearby departing flights are more closely interrelated than those with large time intervals. In addition, a crew is more likely to be substituted by crew members from the same base airport to carry out certain duties than those from other bases. Therefore, we start by ordering flights in chronological order according to their scheduled departure time. Then we define a time window that includes the first

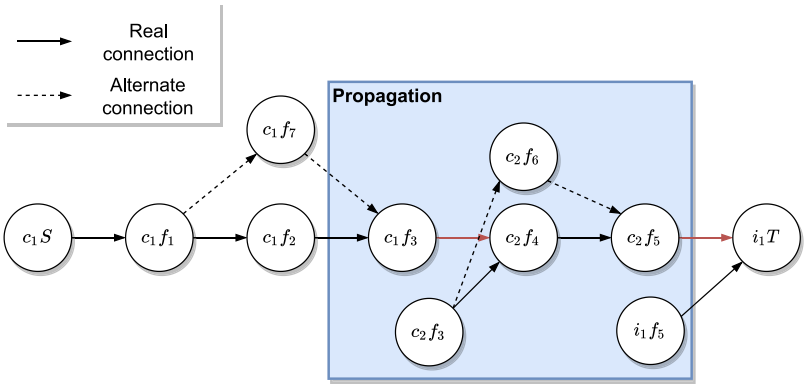
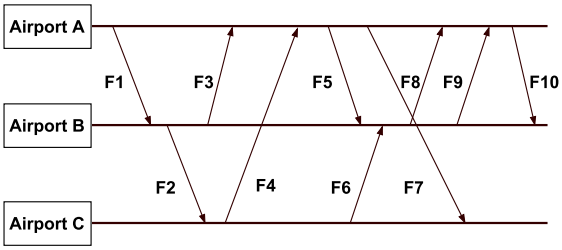


Fig. 3. Example of potential infection routes.



Flight	Departure	Arrival	A/C	Crew	Pax
F1	A (06:00)	B (08:00)	A/C1	C1, C2	P1, P11
F2	B (08:30)	C (10:30)	A/C1	C3, C4	P2, P11
F3	B (10:20)	A (12:20)	A/C2	C1, C2	P3, P14
F4	C (11:00)	A (15:00)	A/C1	C3, C4	P4
F5	A (15:30)	B (17:30)	A/C2	C3, C4	P5
F6	C (16:10)	B (18:10)	A/C3	C5, C6	P6, P12, P13
F7	A (18:00)	C (22:00)	A/C1	C1, C2	P7, P14
F8	B (18:50)	A (20:50)	A/C2	C7, C8	P8, P12
F9	B (21:20)	A (23:20)	A/C3	C5, C6	P9, P13
F10	A (23:50)	B (01:50)	A/C3	C7, C8	P10

Fig. 4. The time-space network and schedules for the 10-flight example.

$|F|/3$ flights as the first neighborhood. The next neighborhood is generated by moving the time window to $[|F|/6, |F|/2)$ with a $|F|/6$ rolling horizon. Therefore, a set of overlapping neighborhoods is generated and a trade-off is reached between solution time and proximity to global optimality. Similarly, crew-related neighborhoods are defined by first aggregating different crew members into several groups given their home bases and randomly placing these groups in a list. The time window and a rolling horizon that contain $|C|/3$ and $|C|/6$ crew members are adopted, respectively.

Step 3: Improving the current solution. To solve the reduced problem, we iteratively release the variables y_f^c belonging to different neighborhoods and fix other y_f^c variables to the incumbent integer solution generated in the previous iteration. The branch-and-cut algorithm is then invoked to improve the current solution based on the previously generated lazy constraints. If improvements are not found within the last three iterations or the relative gap between the current solutions (upper bounds to the optimal solutions) is smaller than the specified tolerance value like 0.5%, we terminate the large neighborhood algorithm. If the reduced problem is not solved to optimal, we shrink the time window (in this context, halving the time window length) to generate two new neighborhoods.

3.4. Illustration with a ten-flight example

We illustrate the PAIR model based on a synthetic example with ten flights to highlight the recovery performance of the proposed model concerning entity infection protection in contrast to the ordinary recovery model without optimization on pandemic propagation. Ten flights denoted by *F1* to *F10* serve three airports. As illustrated in Fig. 4, the flight schedule is displayed in a time-space network where the horizontal direction represents increasing time stamps and flights are indicated with arrows pointing from their departure airports to arrival airports. Eight crew members, denoted by *C1* to *C8*, are scheduled to execute flights with different origin and destination airports. Lastly, 13 passenger itineraries including both non-stop and one-stop itineraries are generated with 100 and 20 passenger demand, respectively. Itineraries can be redirected to an alternative if the same OD market is served and the alternative departs later than the original one. For this specific example, we assume that the positive crew member *C3* is not available and *C4* is infectious with a probability of 0.65 due to close contact with *C3* previously during a six-hour medium-haul flight. In addition, the city of Airport C witnesses rapid growth of a new variant, and passenger itineraries *P4*, *P6*, *P12*, *P13* are assumed with initial infection probabilities of 0.05. For simplicity in this example, the minimum connection time among aircraft, crew, and passengers is set to 30 min and other parameters are set according to Section 4.1.3.

3.4.1. Ordinary recovery solution

For a traditional recovery optimization model without considering the infection propagation effects, variables η and related constraints are accordingly excluded, thus leading to a deterministic model that can be solved with the branch-and-bound method.

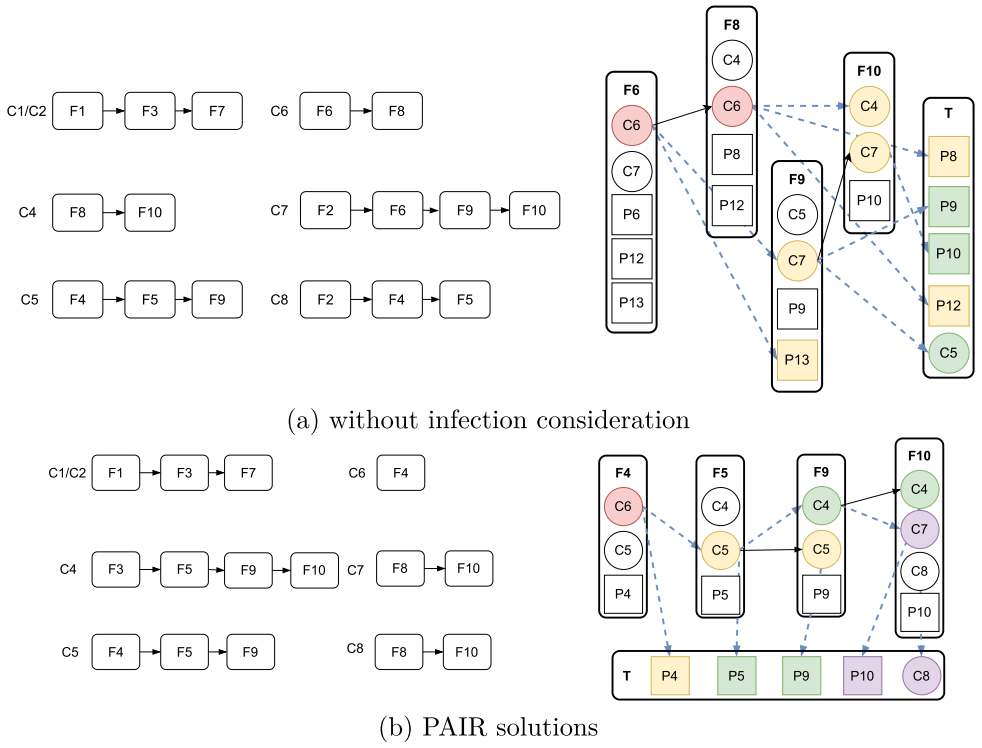


Fig. 5. Recovered crew schedules and epidemic propagation routes from crew *C6*.

The obtained schedules thus cover every flight to reduce the flight cancellation cost. The corresponding crew schedule solution and the possible infection dynamics are depicted in Fig. 5(a) with different colors representing infection orders. To be specific, the newly generated flight plans for crew members *C1, C2* still follow the original schedule while the infectious crew *C6* consecutively executes flights *F6* and *F8* with *C7* and *C8*. The epidemic propagation network graph clearly shows the possible large amount of involved secondary infected crews and passenger itineraries. As close contact and secondary close contact crews/passengers constitute a large proportion of the overall infected entities, the general infection level is relatively high considering that four itineraries from Airport C also incur a large extent of infection. Accordingly, airline recovery without considering epidemic spreading is likely to result in high infection levels.

3.4.2. PAIR based recovery solutions

In contrast, with the proposed PAIR model, the possible infection propagation is penalized. Flights *F2* and *F6* are canceled to keep the flow balance and prevent potential in-flight transmissions incurred by the one-stop passenger itineraries via flight *F6*. Flight *F10* is applied to deadhead one crew member to their original planned overnight airport. Additionally, crew *C6* is assigned to execute only one medium-haul flight *F4* (i.e. carry infectious passenger itinerary *P4*). From the propagation network illustrated in Fig. 5(b), only crew *C5* reaches an expected infection probability of 0.42, and itinerary *P4* is at risk of secondary infection risk. The maximum infection probabilities of other crews and passenger itineraries are 0.042 and 0.003 respectively. This recovery solution differs from the traditionally adopted isolation strategy by utilizing potentially contagious crew members efficiently with contagious passenger itineraries.

In order to validate the use of distributionally robust optimization, the influence of uncertainty degree is first illustrated by enlarging the distance parameter θ from 0.5 to 1.0 where the true distribution of α is assumed to be further away from the empirical distribution than that of $\theta = 0.5$. In this circumstance, infection probabilities can be overestimated to render a conservative solution.

As shown in Fig. 6, cancelling flights *F6*, *F7* and *F8* causes significant disruptions to passenger itineraries *P6–P8*, *P12–P14*. This solution features less infection transmission but also less productivity due to the large-scale perturbations to the existing schedule. As will be explained later, a large θ value does not necessarily lead to small infection solutions as empirical distributions are not well utilized. The related out-of-sample validation to select an appropriate θ for real-world instances is introduced in Section 4.3.2.

4. Experimental results

In this section, a real-world case study is presented to investigate the performance of proposed methodologies and airline recovery reactions based on the network of one carrier in the U.S. COVID-19-related infection disruptions are particularly considered given its unprecedented impact in the recent history of aviation (Sun et al., 2020).

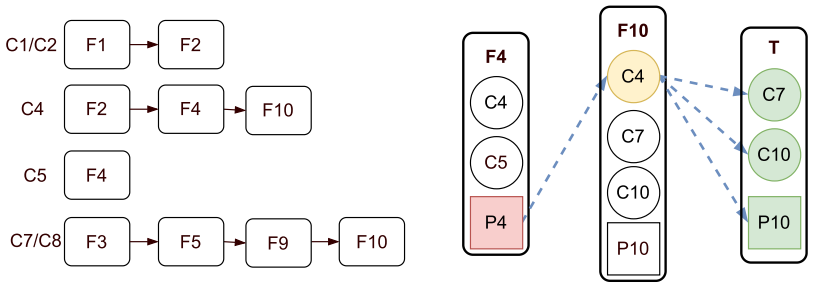


Fig. 6. The recovery solution of PAIR model with $\theta = 1.0$.

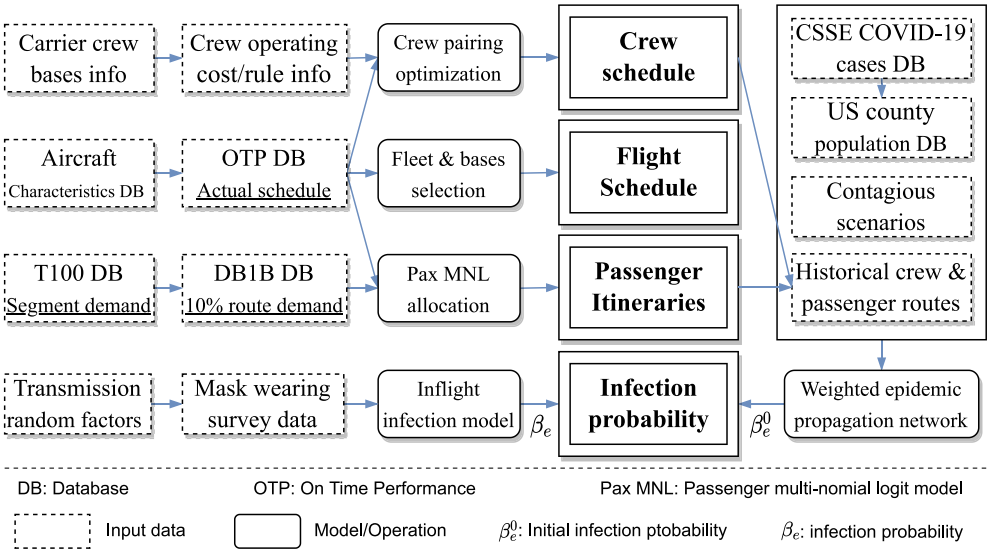


Fig. 7. Data sources and processing diagram.

The solution framework is implemented in C++ and executed on a computer with 2.7 GHz Intel i7-12700H CPU and Ubuntu 22 operating system. All experiments are carried out using a single thread, and ILOG CPLEX 12.10.0.0 is used as the general-purpose solver with its Lazy constraint callback to add cuts dynamically.

Section 4.1 introduces the dataset, parameters, and disruption scenarios. Then, numeric experiments are carried out in Section 4.2 based on the constructed disruption scenarios to validate the performance of the algorithm. Focusing on calibrating the PAIR model, sensitivity analyses for two critical model parameters are carried out in Section 4.3 to reach the balance between operational efficiency and infection risks. Thereafter, the performance of the distributionally robust optimization model is compared with that of a deterministic model and a robust optimization model in Section 4.4 to validate our modeling choice. Finally, Section 4.5 investigates the quantitative impact of crew and passenger infection disruptions.

4.1. Experimental setup

Our computation instances are based on the network of one main legacy carrier in the U.S. which operates its main hub at Dallas/Fort Worth International Airport (IATA code: DFW). To build a series of realistic test instances, multiple open-access data sources and models are adopted in Section 4.1.1 with a complete diagram as shown in Fig. 7. Thereafter, based on multiple disruption types and flight networks, different scenarios are randomly generated in Section 4.1.2 to validate the numerical performance of the PAIR model and solution algorithms. Lastly, the specific parameter values are set and explained in Section 4.1.3.

4.1.1. Input data

To understand the effects of our presented model on real-world airline recovery practices and in order to perform experiments regarding scalability, the U.S. Airline On-Time Performance (OTP) data in November of the year 2021 from the Bureau of Transportation Statistics (BTS) is used to generate realistic computation instances with individual aircraft routes and actual/scheduled

Table 3
Characteristics of the generated instances.

ID	Flights	Crews	Tails	Airports	FT (min)	Pax
N1	81	69	27	35	149	84
N2	123	99	43	45	145	139
N3	159	124	58	50	147	189
N4	192	146	74	52	147	206
N5	211	158	93	52	152	234
N6	230	172	93	52	153	281

time of arrival/departure.[2] Based on the civil registered aircraft type information of the FAA, the weekly flight schedule of B737-800 containing 167 seats from a major legacy carrier is extracted.

Using the aggregated passenger demand dataset *T-100 Domestic Segment* and *Airline Origin and Destination Survey (DB1B)* from BTS, a disaggregation approach from Barnhart et al. (2014) is applied to generate specific passenger itineraries considering factors include departure time, connection time, seating capacity and time zones. Specifically, the carrier route demand from DB1B is extracted and calibrated according to the segment demand T-100 dataset. Potential passenger itineraries consisting of one or multiple flight legs are generated from the OTP flight schedule. Subsequently, a multinomial logit model is used to allocate passengers' route demand to different itineraries allowing for choice probabilities and flight capacity.

Since crew schedule data is not publicly available, a four-day crew pairing solution is generated accordingly in our study using the optimization model and branch-and-dive algorithm from Saddoune et al. (2013). A number of realistic crew duty and pairing regulations are considered in this context including the maximum pairing duration, briefing/debriefing periods, the minimum rest time, and so forth. In addition, a set of crew-related costs are incorporated in terms of credit flying time, waiting, resting, and deadhead. Additional crew hiring costs are added to derive efficient pairings originating from 10 fixed crew bases.

For infection-related datasets, U.S. county mask-wearing survey data[3] is adopted to get the statistics of regional-varying mask-wearing ratios. Concerning the evaluation of the initial infection probabilities, county-based CSSE COVID-19 confirmed case data (Dong et al., 2020) is used to compute the existing case ratio of passenger itineraries which is the newly reported cases of seven days divided by the county's population. To evaluate the initially crew infection probabilities, a complete epidemic propagation network is constructed based on the executed pairings of involved crew members and passenger itineraries. Consequently, the initial infection probability of every crew member is obtained with the aforementioned graph search algorithm.

4.1.2. Disruption scenarios

Based on the input data, six problem instances are constructed with varying sizes and complexities whose aggregated characteristics are summarized in Table 3. For each instance, we report the number of flights, aircraft, average flight time in minutes *FTime*, and the number of passenger itineraries *Pax*. By covering small and medium-sized airline fleets, these generated instances help evaluate the performance of the proposed model and algorithms under different real-world operating conditions. The flight networks of all instances are depicted in Fig. 8, enabling a visual comparison of the differences between each network. From this figure, the six instances from N1 to N6 represent different airline marketing patterns with one or more hub airports (e.g. DFW and CLT). With the increase in network size and operating entity number, more recovery opportunities are likely to be provided.

Concerning the construction of disruption scenarios as test cases, we consider four kinds of disruptions including (1) airport closure, (2) aircraft-on-grounds, (3) delayed crew ready time, and (4) aircrews tested positive for certain diseases. The first three disruptions have been widely studied in the past research and regional infection levels are set according to the aforementioned county infection dataset. Three scenarios **SC1**, **SC2**, and **SC3** for each network instance are randomly generated. To be specific, the hub airport DFW is assumed to be closed for a time period selected from a uniform distribution within three to six hours. One aircraft is randomly selected to undergo mechanical failure for 1–12 h. Concerning the delayed crew ready time, 1%–10% crew members are selected to be unavailable for 20 min–12 h. Lastly, for infection-related disruptions, 1%–10% crew members are also randomly selected to be confirmed cases and one flight on the previous day is assumed to report 1–10 COVID-positive passengers. After that, based on the historical crew schedules along with passenger itineraries of the previous three days, an epidemic propagation network is built and edges from these confirmed entities to the dummy sink nodes are assigned with an initial value $\ln 1 = 0$. Latter, an acyclic shortest path algorithm is run on the **reverse** graph. In this way, the infection probabilities of crew members before the day of operation can be obtained by checking the distance at the source vertex of every crew member.

4.1.3. Parameter settings

Multiple parameters are involved in this study to capture the joint economic impact of recovery decisions. Specifically, following Arıkan et al. (2017), we set the flight delay cost tc_f^2 as $\sum_{i \in I_f} D_i \cdot 1.0242$ to approximate the passenger delays which is more computationally efficient than using the individual level delay. The unassigned passenger cost tc_i^4 is set to \$457.8 with respect to the average ticket cancellation cost. Correspondingly, tc_f^1 is set to zero to avoid re-penalizing the flight cancellation decisions. The crew deadhead cost tc_f^3 is set to \$1000 as defined in Petersen et al. (2012).

[2] https://www.transtats.bts.gov/ontime/.

[3] The New York Times. (2021). Coronavirus (Covid-19) Data in the United States. Retrieved 2022-08-10, from https://github.com/nytimes/covid-19-data.

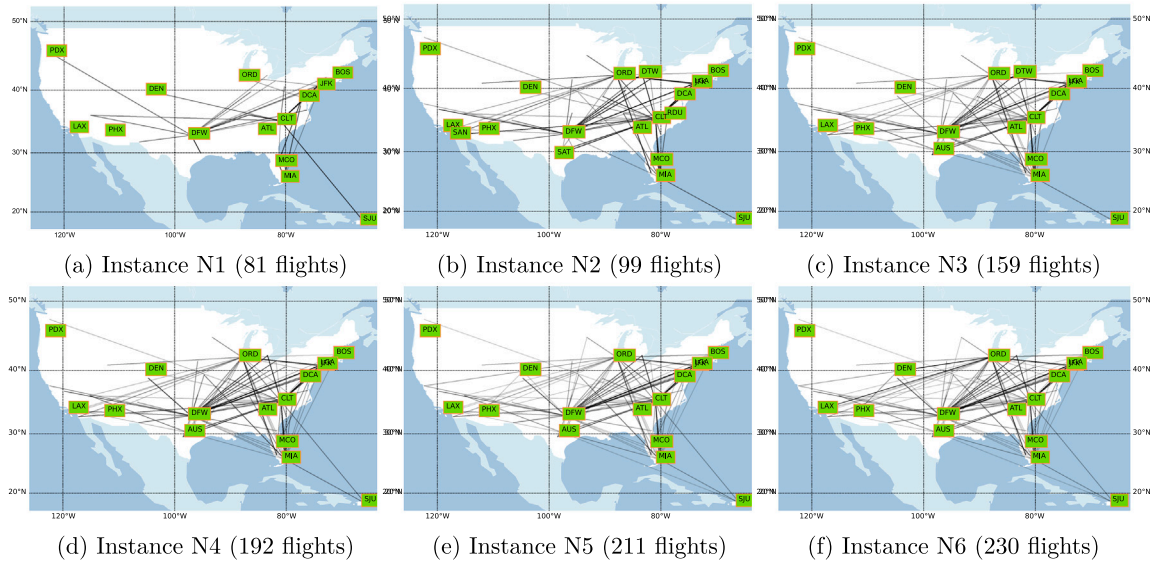


Fig. 8. Visualization of the generated instances with airport locations and flight links. Different thickness of the lines indicate flight frequency.

As for infection-related cost, crew infection incidence is likely to lead to subsequent flight delays and the usage of reserve crew members at a premium pay (Shebalov and Klabjan, 2006). Although the cost of using reserved crew members is set to be 25 times higher than unit delay cost (Abdelghany et al., 2004), we do not include this value as it overlooks the possible crew absence and flight cancellation during a potential recovery period once infected. Instead, the tc_c^5 is derived as the flight cancellation cost $tc_c^5 = \$20,000$ from Arıkan et al. (2017) and other possible values with their effects on operating cost/infection propagation are discussed further in Section 4.3. For passenger infection social cost, the airline compensation fare ($42.7/h) for layovers from Luttmann (2019) is adopted for a two-week recovery period[4] $tc_i^5 = D_i \cdot 14,347$ to serve as the passenger value of time.

While multiple parameters of the in-flight transmission risk evaluation model in Appendix A are set to fixed distributions, the variation of regional mask-wearing rates p_{mask} is taken into account in this study by adjusting the mean value with a multiplier derived from the aforementioned mask-wearing survey data (including frequency ratios higher or equal to **sometimes**). Moreover, compared to the widely available city/country COVID-19 case dataset, in-flight transmission data including the seat position of index/secondary patients, masking condition, flight duration, and individual information (e.g. age, region) is relatively hard to access (Kelly et al., 2021), while airlines are expected to access and collect the corresponding data more easily especially for cabin crew-related data in order to calibrate the prediction model more precisely. Because the effective reproduction number of the Omicron variant is reported to be 3.19 times greater than that of Delta (Ito et al., 2022) and the reported cases used in Barnett and Fleming (2022) date back to 2020, we adjust the beta distribution so that $\pi_0 \sim Beta(3.19, 520)$. Furthermore, parameter $k_1 = 0.814$ for flight duration effects $\frac{FT_f}{60}^{k_1}$ is estimated using non-linear least square method.

Finally, regarding the selection of parameters for distributionally robust optimization, the infection probability samples N is set to 50 to serve as empirical distributions. The ambiguity distance θ is set to 0.5 which serves as a reasonable gap between the empirical distribution and the true one as indicated by the experimental results later.

4.2. Numerical results

Experiments are carried out for 18 scenarios that combine different networks and disruption scenarios using the improved branch-and-cut solution algorithm. The connection threshold φ used in the construction of the flight connection network is set to be 50 min and the maximum number of edges allowed in an individual connection network using the network size control algorithm presented in Arıkan et al. (2017) is set to be 500, where both parameters are validated to achieve good performance in terms of convergence and solution quality. The maximum solution time is set to 30 min.

In Table 4, statistics including model constraints *Cons*, variables *Vars*, the objective value, solution time and the generated cuts of the branch-and-cut and large neighborhood search algorithms are reported for each instance *Inst* and disruption scenario. The relative gaps between two algorithms $\frac{Obj_2 - Obj_1}{Obj_2}$ are reported in column *Gap*.

As shown in the table, the model size increases steadily with the instance size and the solution time for large instances is observed to follow a surge increase trend. The branch-and-cut algorithm soon faces computational intractability from instance N3 to N6. On

[4] Active illness can last one to two weeks for moderate coronavirus disease https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/diagnosed-with-covid-19-what-to-expect.

Table 4
Computation results of PAIR model using branch-and-cut and large neighborhood search.

#	Inst	Disrupt	Cons	Vars	Branch-and-cut			Large neighborhood search			Gap (%)
					Objective	Time (s)	Cuts	Objective	Time (s)	Cuts	
1		SC1	5,373	4,347	1,364,136.25	1.69	259	1,364,136.25	1.84	259	0.00
2	N1	SC2	5,402	4,347	1,515,235.19	0.57	174	1,515,235.19	0.64	174	0.00
3		SC3	5,383	4,349	1,542,879.54	1.89	286	1,542,879.54	2.08	286	0.00
4		SC1	12,240	10,089	1,448,413.56	1,800.00	769	1,448,417.28	172.84	873	0.00
5	N2	SC2	12,669	10,489	1,026,094.65	144.66	629	1,026,094.66	130.30	776	0.00
6		SC3	11,687	9,593	1,436,958.78	1,800.00	570	1,436,966.28	165.07	637	0.00
7		SC1	18,864	15,344	1,737,370.68	1,800.00	985	1,758,943.38	187.06	936	1.24
8	N3	SC2	19,433	15,970	1,450,262.27	1,800.00	1,160	1,449,935.74	421.61	1,268	−0.02
9		SC3	19,596	16,070	1,657,257.41	1,800.00	1,375	1,671,821.05	335.65	1,455	0.88
10		SC1	30,807	25,820	1,841,288.32	1,800.00	676	1,858,961.54	498.54	921	0.96
11	N4	SC2	30,832	25,820	1,834,082.75	1,800.00	774	1,849,601.55	629.82	1,212	0.85
12		SC3	26,316	21,713	2,589,018.32	1,800.00	775	2,589,018.29	265.39	734	0.00
13		SC1	36,997	31,073	2,798,900.43	1,800.00	1,360	2,800,866.87	460.52	1,924	0.07
14	N5	SC2	36,920	31,073	2,084,830.84	1,800.00	602	2,094,397.87	443.41	807	0.46
15		SC3	34,607	28,974	2,610,440.84	1,800.00	1,342	2,606,094.51	874.44	1,481	−0.17
16		SC1	36,331	30,267	3,283,965.87	1,800.00	367	3,323,532.87	504.50	347	1.20
17	N6	SC2	43,169	36,377	2,445,783.73	1,800.00	692	2,486,347.47	840.17	1,964	1.66
18		SC3	43,098	36,377	2,237,511.61	1,800.00	2,085	2,215,832.83	937.71	2,327	−0.97

the other hand, the developed large neighborhood search algorithm is able to solve all instances within 15 min within a 1.7% relative gap. Such superiority can be partially indicated by the number of generated cuts where a larger amount of cuts are added to each reduced problem in the large neighborhood search algorithm due to the reduced complexity when solving the reduced problems using the fix-and-release operator. Overall, our proposed model and the corresponding solution methodology help to solve real-time integrated airline disruption management problems efficiently.

4.3. Sensitivity analyses of critical parameters

While the economic cost of flight delay and cancellation is widely analyzed in the literature, it is less clear for the crew infection as it may cause a severe shortage of eligible crew members and an increase in flight cancellations. Besides, the adoption of parameter θ, which controls the radius of the ball around the empirical distribution, may pose different impacts on the actual infection risks. When $\theta \to 0$, solutions that ignore distributional ambiguity are obtained with overfitting risk, when increasing θ to a relatively large number, the information of empirical distributions is gradually overlooked. Hence, two sensitivity analyses are carried out to investigate the reasonable range of tc_c^5 and θ.

4.3.1. Experiments for parameter tc_c^5

To investigate the reasonable bounds of tc_c^5, the maximum possible cost that leads to no crew infection is first calculated as the unassigned passenger cost of four flights $4tc_i^4 \cdot S$ given $MLN_c = 4$ along with a flight load factor of 1. Because this value is rather large and may lead to over-conservative solutions, $tc_c^5 = 80,000 \approx tc_i^4 \cdot S$ is used as the base reference and experiments are conducted to vary the fraction of tc_c^5 base value from zero to 2.0 for network instance N5 with delayed crew ready time and infection disruptions as an example.

The computational statistics are demonstrated in Fig. 9 which reports the normalized values of flight delays and cancellations with expected entity infection number. The normalized value is computed by dividing the flight cancellation number (ranging from 17–25) and the total flight delay time (ranging from 216–403 min) by their maximum values, respectively. From the figure, it is apparent that prioritizing crew infection in optimization helps to cut down the overall transmission probabilities of the derived recovery schedules significantly in contrast to the $Fraction = 0$ case where 44 crew and passengers are likely to be operated with risk. In addition, while the infection level witnesses a rapid decrease when the fraction value is no more than 0.5, further increasing tc_c^5 only provides marginal improvement with much higher flight delays. Hence, the reasonable range for tc_c^5 can be $[0.1 - 0.5]$ and the default cost parameter $tc_c^5 = 20{,}000$ effectively reduces the in-flight epidemic propagation with a relatively high operational performance (avoid high delay and cancellation levels).

Further visualizing four cases in Fig. 10 gives a clear picture of the effectiveness of the proposed PAIR model. In this figure, red lines indicate flights that ultimately lead to the infection propagation and the brightness is correlated with the maximum resulting expected infection likelihood of any crew/itinerary via that airport pair, whilst the brightness of black lines indicates the flight frequency between the specified two airports. A significant amount of flights are therefore protected and limited infection is restricted to a few main hub airports (i.e. DFW, ORD, CLT). On the contrary, disregarding disease transmission (Fig. 10(a)) poses remarkable side effects on more than ten important airports.

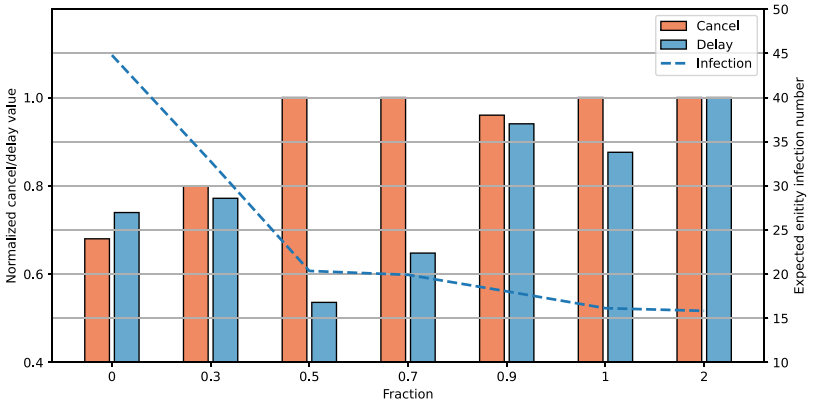
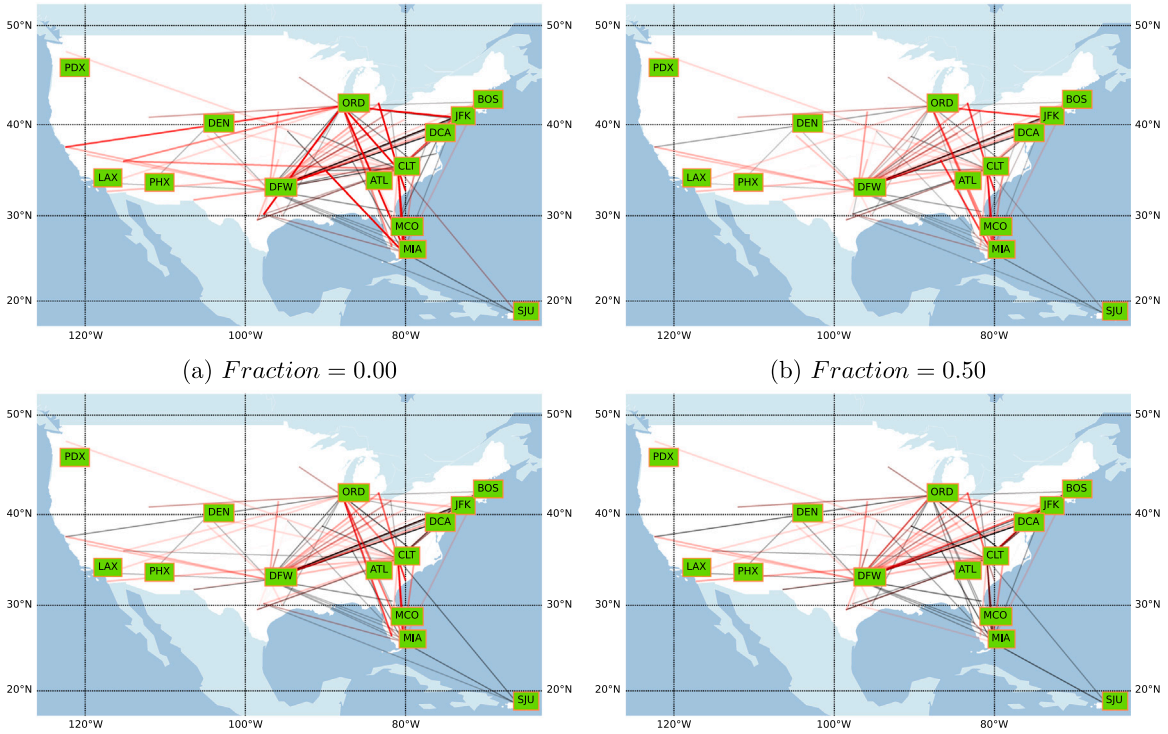


Fig. 9. Influence of crew infection cost on airline operation and entity infection number.



(a) $Fraction = 0.00$ (b) $Fraction = 0.50$

(c) $Fraction = 1.00$ (d) $Fraction = 2.00$

Fig. 10. Epidemic propagation trajectories for different infection cost with PAIR model.

4.3.2. Experiments for parameter θ

One of the main characteristics of the presented model is the adoption of distributionally robust optimization instead of stochastic programming which is known to cause potential overfitting solutions based on the empirical distribution. Accordingly, a set of θ values is tested and compared on the aforementioned disruption instance where 1000 infection rates $\hat{\boldsymbol{\alpha}}$ samples for each crew member and passenger itinerary are generated as the out-of-sample test data.

Generally, the case of $\theta = 0$ ignores the ambiguity of assumed distributions and is reduced to the sample average approximation method. The out-of-sample experiment results are shown in Table 5 and the first six columns except for the first one report the parameter value, number of canceled flights, the summation of flight delays, number of invalid passenger itineraries due to cancellation or short connection time *Invalid ITIN*, and the expected infection number, respectively. Whilst the last four columns depict the distribution of infected crews and passenger itineraries.

The table shows that, in general, the out-of-sample results of the expected infection number decrease after an initial increasing trend. When θ is set to 0.4, the overall expected infection number outperforms others with regard to the sampled data but is likely

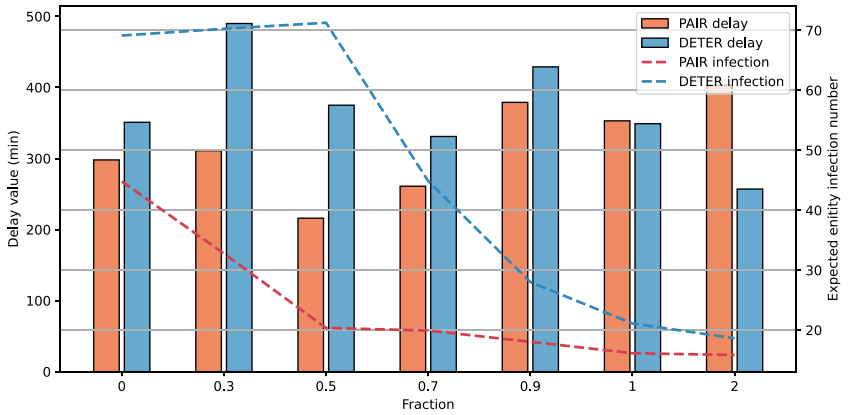


Fig. 11. Flight delays and entity infection with DETER and PAIR models.

Table 5
Sensitivity analysis statistics for distance parameter θ.

θ	Cancel	Delay (min)	Invalid ITIN	Infect No.	Infection probability distribution (%)			
					[0.0, 0.2)	[0.2, 0.4)	[0.4, 0.6)	[0.6, 1]
0.0	18	261	23	74.33	84.58	6.91	0.01	7.62
0.2	17	298	22	67.29	83.29	7.50	0.02	9.19
0.4	20	321	25	38.94	90.59	5.29	0.00	4.14
0.6	18	331	22	61.87	86.15	6.53	0.00	7.32
0.8	18	497	23	68.21	85.79	6.76	0.01	7.45
1.0	19	375	23	74.59	85.78	6.85	0.00	7.37

to cause high operation costs as indicated by the delay and cancellation numbers. In this case, the number of entities with moderate infection probabilities is relatively small. The results show that incorporating ambiguity over empirical distribution helps to improve the out-of-sample performances of the selected crew infection risk evaluation criteria and the default value $\theta = 0.5$ is also a good candidate to avoid overfitting.

4.4. Assessment of model performance

The previous experimental results have highlighted the importance of considering in-flight epidemic propagation, more interestingly, we aim to investigate the necessity of our uncertainty modeling using the distributionally robust optimization approach against other possible alternatives. In particular, a deterministic recovery model (DETER) and a commonly used robust optimization (RO) model are developed and validated to illustrate the corresponding effects on operational efficiency and disease dissemination.

4.4.1. Experiments for DETER model

To concretize the commonly applied airline strategies for in-flight disease transmission, we design the deterministic DETER model that penalizes contacts between risky crew members and other onboard entities. This isolation-alike strategy is commonly adopted in practice to hinder epidemic spreading (Cristian Ionut Panait, 2020). An additional variable $p_{c,e} \in \{0,1\}$ is added for each pair of crew and entity to indicate whether or not crew c and entity e have served one common flight. The additional model formulation is presented as follows:

$$\eta_e \geq \beta_e^0 \bar{\beta} p_{c,e} \qquad \forall c \in C, \zeta_{c,e} = 1 \tag{40}$$

$$p_{c_1,c_2} \geq y_f^{c_1} + y_f^{c_2} - 1 \qquad \forall f \in F \tag{41}$$

$$p_{c,i} \geq y_f^c + s_i - 1 \qquad \forall f \in F, i \in I_f \tag{42}$$

Here, Constraints (40) compute the infection risk value between risky crew member c and entity e depending on the risk level β_c^0, average infection probability $\bar{\beta}$ and binary parameter $\zeta_{c,e}$ which equals 1 if $\beta_c^0 > \beta_e^0$. Constraints (41) and (42) ensure the lower bound on variables p by checking every possible combination of entities and flights for crews and passengers, respectively. Due to the fact that connection network size can be controlled for every individual crew member, the cardinality of the new formulation remains reasonable. For the minimization problem, the optimal values of variables p are obtained without restricting the upper bounds.

Based on the same computation configuration used in Section 4.3.1, a set of experiments with varying crew infection costs are carried out to derive the comparative performance of DETER model against the PAIR model in Fig. 11. Although the deterministic

Table 6
Computation results for different parameters of the RO model.

γ	$\lvert B_l \rvert$	Cancel	Delay (min)	Invalid ITIN	Infect No.	Infection probability distribution (%)			
						[0.0, 0.2)	[0.2, 0.4)	[0.4, 0.6)	[0.6, 1]
0.6	2.0	22	393	27	25.19	93.26	5.78	0.00	0.00
	4.0	29	282	35	16.63	95.96	4.04	0.00	0.00
	8.0	31	416	36	15.28	97.00	3.00	0.00	0.00
0.9	2.0	27	345	32	22.87	94.74	3.94	0.01	1.32
	4.0	33	329	40	13.51	96.97	3.03	0.00	0.00
	8.0	31	344	37	15.36	95.60	4.40	0.00	0.00

recovery model explicitly discourages the combination of risky crew members and other entities. In general, it cannot provide efficient and economical solutions consistently since its ignorance of the detailed epidemic spreading and passenger-induced propagation. When $Fraction \leq 0.5$, the disturbed crew schedule even slightly degrades performance since infection probabilities and propagation are not well modeled. When $Fraction = 0.25(tc_c^5 = 20,000)$, the expected infection number of the PAIR model is over 45% lower than that of DETER model. Further enlarging the cost parameter will gradually reduce the difference in infection number between the two models as most contagious crews are not used. Besides, the DETER model features comparatively high delay rates (due to reduced crew utilization rates) and low cancellation rates when $Fraction \leq 1$. Overall, purely isolation-based crew operations may not fully utilize airline resources, while appropriate uncertainty modeling of epidemic transmission, in conjunction with other recovery decisions will enable stable airline operations under public health issues.

4.4.2. Experiments for RO model

The distribution-free RO model is developed to derive robust recovery solutions where the worst-case infection cost of a route r based on the widely used budget uncertainty set is defined as follows.

$$\max \sum_{(e_1,e_2)\in r} \alpha_{e_1,e_2} \tag{43}$$

$$\alpha_{e_1,e_2} \in [\alpha^L_{e_1,e_2}, \alpha^U_{e_1,e_2}] \qquad \forall (e_1,e_2) \in r \tag{44}$$

$$\sum_{(e_1,e_2)\in B_l} \alpha_{e_1,e_2} \leq b_l \qquad \forall l \in L \tag{45}$$

The uncertainty set is made up of $|L|$ budget constraints and requires that the cumulative log transmission rates for passengers/crews in the set B_l should not exceed an upper bound b_l. Similar to Gounaris et al. (2013), crew and itineraries are classified into different categories based on their geographical locations. The sample data is used to calculate the $\boldsymbol{\alpha^L}$ and $\boldsymbol{\alpha^U}$, as well as the mean value $\boldsymbol{\alpha^M}$ and variance $\boldsymbol{\sigma}$. Then based on the Lyapunov central limit theorem from Subramanyam et al. (2020), b_l is approximated as $\sum_{(e_1,e_2)\in B_l} \alpha^M_{e_1,e_2} + \Phi^{-1}(\gamma) \sum_{(e_1,e_2)\in B_l} \sigma^2_{e_1,e_2}$ where $\Phi^{-1}(\gamma)$ is the inverse cumulative distribution function on probability γ of the standard normal distribution.

The closed-form expression of model (43)–(45) can be deduced using linear duality theorem as $\sum_{(e_1,e_2)\in r} \alpha^U_{e_1,e_2} - \sum_{l\in L} \max\{0, \sum_{(e_1,e_2)\in r\cap B_l} (\alpha^U_{e_1,e_2} - \alpha^L_{e_1,e_2}) - (b_l - \sum_{(e_1,e_2)\in B_l} \alpha^L_{e_1,e_2})\}$ and replaces Eq. (35) used in the multi-label shortest path algorithm. In this section, two γ values and three cardinality values of the budget uncertainty sets $|B_l|$ are chosen.

Smaller γ directly reduces the value of b_l, whereas reducing the value of $|B_l|$ restricts the uncertainty set with more information. To understand the impact of the robust optimization approach with the two key parameters, the RO model is implemented and optimized as shown in Table 6 based on the out-of-sample test data presented in Section 4.3.2.

From the results, increasing $|B_l|$ and γ generally helps to derive more robust solutions against uncertain infection risks due to the increase in b_l. Although the RO approach achieves schedules with lower infection risk than the DRO approach, it has a high level of flight cancellations and invalid itineraries, indicating inefficiency and low productivity. For instance, when $\gamma = 0.6, |B_l| = 8$, the corresponding infection number is close to the PAIR model with $Fraction = 2.0, \theta = 0.5$ from Fig. 9 but with a substantially higher level of flight cancellation (25 versus 31).

To summarize, extensive contrasting experiments validate the superiority of the PAIR model over sample average approximation ($\theta = 0$), deterministic DETER model (directly reducing individual contacts) and budget uncertainty set-based robust optimization (over-conservative). The model is also able to efficiently trade off between in-flight pandemic transmission and operational performance in relation to a family of probability distributions of uncertain transmission rates.

4.5. Quantifying the influence of infection disruptions

From an operational point of view, protecting crews and passengers from epidemic spreading entails postponing departure/arrival time, swapping aircraft/crews for specific flights, and reassigning passenger itineraries. Thereby, it is of interest for airline operators to understand the comparative importance of crew/passenger infection and the relative influence of epidemic infection on other disruption scenarios. Experiments are conducted accordingly to investigate these relationships.

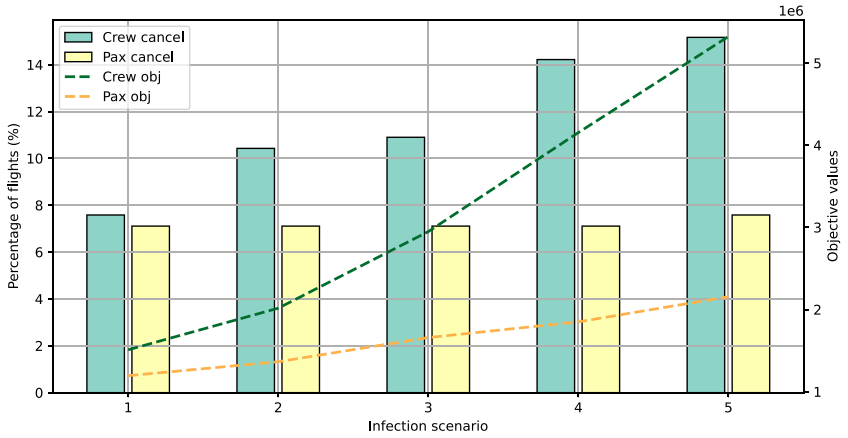


Fig. 12. The comparative results of disruptions with crew and passenger infection.

4.5.1. Comparative importance of crew and passenger infection

Based on the same *N5* test instance with randomly mixed disruptions of airport closures and aircraft-on-ground, different numbers of crews and passengers are assumed to be infected, respectively. To ensure a fair comparison, when n crew members are sampled as infected individuals in one disruption scenario, $\frac{tc_i^5}{D_i \cdot tc_c^5} n$ passengers from different itineraries are randomly selected to build the contrasting scenario because $tc_i^5 = D_i \cdot 14,374$. In this context, the initial infection probability β^0 for every crew member is set to one whilst that for every itinerary is the quotient of risky passenger number and itinerary demand. In Fig. 12, five disruption scenarios with varying numbers of infected crews (passengers) are examined: 5 (7), 10 (14), 20 (28), 30 (42), and 40 (56). Bars represent percentages of canceled flights (the base value without infection disruption is 6.63%) and dashed lines represent model objective function values.

From the results, it is found that although the increasing number of infectious entities leads to rising objective function values, the impact of the two entity types on recovery performance is heterogeneous. 56 infected passengers cause only one additional flight to be canceled, whereas crew resources have a greater impact, with 8.53% additional flights canceled when 40 crew members are infected. This is because aircrews typically possess opportunities to contact more people on multiple flights when executing their duties and the higher operating cost (coupled with aircraft routes) than the passenger unassigned cost.

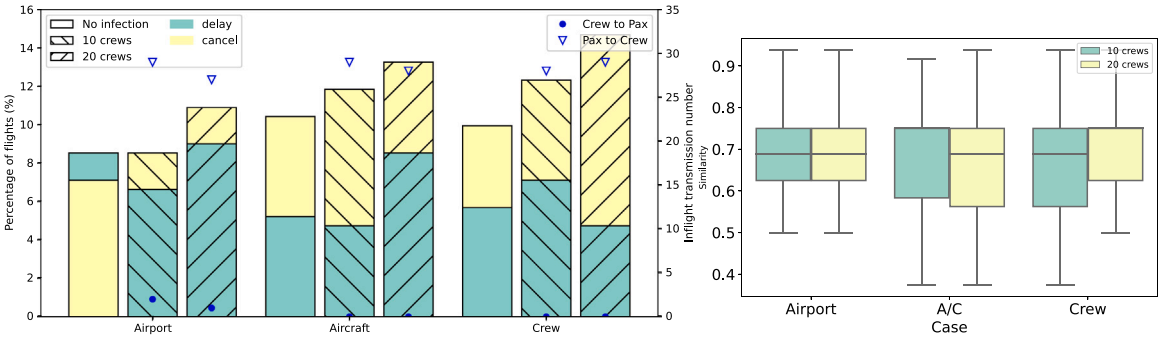
4.5.2. Influence of epidemic infection on other disruptions

To answer the question: How much extra effort do airlines need to obtain an infection-safe recovery solution except for other commonly encountered disruptions? Nine scenarios based on network N5 are generated by combining the three base disruption types with three crew infection disruptions to provide explicit comparisons since crew resources have been shown to be more influential than passengers. The three crew infections consist of no infection, 10 (6.33%) infected crews, and 20 (12.66%) infected crews, while passenger infection levels are set according to the real-world data in Section 4.1.1. The three base disruptions are listed as follows.

1. Hub airport closure for three hours.
2. Five aircraft-on-grounds for 1–12 h.
3. Ten crew members undergo delayed ready time for 20 min–12 h.

The detailed computation statistics are summarized in Table 7. Except for the previously defined evaluation criteria, two scores *Sim A/C* and *Sim Crew* represent the similarity between the crew/aircraft routes of the base case and those of the case with initial crew infection disruptions. Normalized compression distance (NCD) from Xu et al. (2021) is used to calculate these scores. An aircraft or crew route is defined as the combination of actual departure hours, departure airports, arrival hours, and arrival airports. By selecting a general compressor Z to build the suffix tree, the NCD for two routes r_1, r_2 is computed as $\frac{Z(r_1 r_2) - \min\{Z(r_1), Z(r_2)\}}{\max\{Z(r_1), Z(r_2)\}}$. Then the NCD value for route r_1 is taken as $\max_{r_2}\{NCD_{r_1,r_2}\}$ to capture the case when the same routes are taken by different aircraft/crews.

From the results, crew infection is not found to be quite destructive to the original recovery schedules for other commonly solved disruptions. Moderate additional flight delays and cancellations (13%–53% increase) are needed to eliminate crew infection propagation. While delay and cancellation percentage increments are visualized in Fig. 13(a), it is observed that only case 6 involves high flight delay increments (60% higher than the base), since the recovery solutions try to keep the original aircraft sequence to avoid extensive crew swapping and the corresponding high infection risks. A high level of cancellations is observed for a delayed crew ready time disruption (case 9) since there is an urgent shortage of available crew resources. In addition, infection event numbers from crews to passengers and from passengers to crews are also recorded with points in Fig. 13(a). The value of *Pax to Crew* indicates the severity of delayed crew ready time in terms of infection since high crew utilization rates are expected for such conditions. Due to the relatively larger population size, more disease spreadings from passenger itineraries to crews are found in these experiments.



(a) Increments of delays, cancellation and infection (b) Crew similarity distribution (NCD<1)

Fig. 13. Comparison of recovery solutions for nine disruption scenarios.

Table 7
Comparison between infection and other disruptions.

#	Target	Infection	Cancel	Delay (min)	Inv Pax	Sim A/C	Sim Crew
1		–	15	502	19	1.00	1.00
2	Airport	10 crews	18	441	24	0.87	0.86
3		20 crews	23	517	29	0.86	0.84
4		–	22	320	26	1.00	1.00
5	Aircraft	10 crews	25	312	30	0.85	0.85
6		20 crews	28	512	33	0.83	0.83
7		–	21	425	28	1.00	1.00
8	Crew	10 crews	26	495	33	0.83	0.85
9		20 crews	31	329	38	0.84	0.86

The average similarity also provides evidence that infection-related resolutions feature reasonable deviations from the original schedule. When the similarity distribution is visualized in Fig. 13(b), the more severe case 9 is found to resemble higher similarity with the base case than the less severe case 8. Such a phenomenon is attributed to the reduced solution spaces restricted by the increasing number of infection crew members. To conclude, jointly optimizing infection disruptions with other disruptions does not pose a big interference in the current disruption management practices. Airlines do need to control the overall infection level at an early stage so as to increase the robustness against potential operational inefficiencies caused by epidemic propagation.

5. Conclusion

This study proposes a novel approach to integrated airline recovery under in-flight pandemic transmission risks. The PAIR model hybridizes the aircraft recovery, crew recovery, and passenger reassignment decisions with the uncertainty of epidemic transmission probability captured through a Wasserstein distance-based ambiguity set. To efficiently solve the model, we present a branch-and-cut algorithm combined with the large neighborhood search heuristic to iteratively add infection cost-related cuts based on the established epidemic propagation network. Based on a real-world case study that involves the COVID-19 pandemic, the studied PAIR model is able to be solved with reasonable computation effort and the experimental results illustrate superior performances of the PAIR model with respect to operating costs and in-flight infection over both a deterministic approach or a robust optimization approach. In practical terms, the proposed model and algorithms provide innovative schedule recovery insights and decision support for the airline operations control center to manage public health events along with other operational disruptions. This approach enables airlines to safeguard the health of onboard crews and passengers whilst mitigating the possible crew shortages caused by illness and delayed ready time. Although this may result in slight increases in flight cancellation/delay rates (no more than 4%), the benefits of ensuring the health and safety of passengers and crews far outweigh these costs. Therefore, the proposed approach has significant practical implications for airlines in terms of minimizing economic losses and enhancing their ability to respond to and recover from disruptions, ultimately contributing to a more resilient and sustainable aviation industry.

To the best of our knowledge, our work is the first to optimize the airline disruption management problem under disease transmissions. There are, however, some limitations to be noted. First, although suggestions are provided regarding the estimation of infection probabilities and transmission rates from the prediction model, extensive studies are still needed by airlines and researchers to identify an effective and reliable approach to dealing with aircrew-specific data with certain validations. Second, we restrict our integrated recovery model to incorporate in-flight transmissions due to the complex infection possibilities incurred by other sources (e.g. airport check-in and screening). This factor ought to be considered in the future along with local anti-epidemic measures to enhance the operation stability.

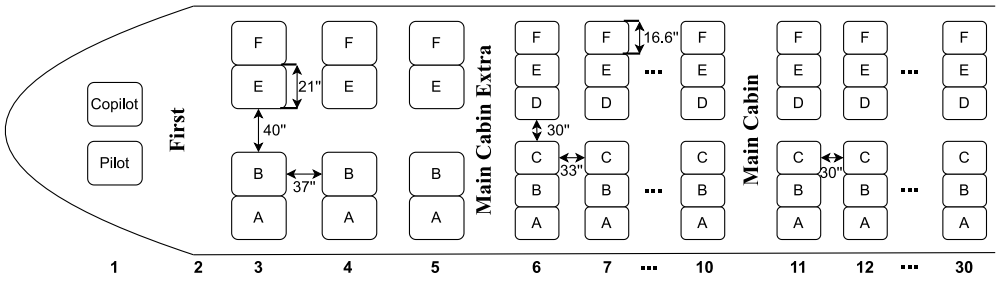


Fig. A.14. The seat map of the B738 fleet.

We summarize a few future research directions based on our work. Improved infection process modeling and evaluation criteria are expected to enhance the model's applicability in real airline practices. Other risk-based decision criteria like the conditional value at risk are promising to be incorporated into model operators' customized tolerance regarding crew infection rates. Furthermore, the model can be extended to solve proactive airline schedule planning problems that jointly take pandemic-induced demand uncertainty into account. More robust solutions against epidemic spreading and demand fluctuation shall benefit airlines with improved revenues over a relatively long planning horizon.

CRediT authorship contribution statement

Yifan Xu: Conceptualization, Methodology, Software, Writing – original draft. **Sebastian Wandelt:** Conceptualization, Validation, Writing – reviewing & editing. **Xiaoqian Sun:** Conceptualization, Validation, Writing – reviewing & editing.

Acknowledgments

This study is supported by the Research Fund from National Natural Science Foundation of China (Grants No. U2233214 and Grant No. 62250710166). We would like to appreciate the suggestions provided by discussants and committee members from the Symposium on Aviation Research (SOAR) conference.

Appendix A. Infection transmission risk evaluation

As a special use case, the COVID-19 pandemic-related in-flight transmission risk is evaluated in this context. To accurately model the uncertain impact factors including flight duration, load factor, and close contagious risk, the infection probabilistic model of Barnett and Fleming (2022) is modified and applied here. The infection probability $P_{e_1e_2}$ from entity (crew/passenger) e_1 to entity e_2 can be computed as $\tau^f_{e_1}\beta_{e_1e_2}$ where the constant parameter $\tau^f_{e_1} = N_{e_1}\frac{FT_f}{60}^{k_1}$ represent the impact of infector number of entity i and the elapsed time (min) of flight f on infection with scale k_1. The random infection rate $\beta_{e_1e_2}$ from entity e_1 to e_2 with a single infector on a 60-min flight is calculated as follows.

$$V^{e_1}_{S_1,S_2} = \pi^{e_1}_0 e^{-\omega d}(1-\lambda)^R(1-p_{masks}) \tag{A.1}$$

$$U^{e_1e_2}_{S_1} = \frac{1}{N-1}\sum_{S_2\in S_{e_2}, S_2\neq S_1}(1-(1-V^{e_1}_{S_1,S_2})^{60}) \tag{A.2}$$

$$\beta_{e_1e_2} = \frac{1}{N}\sum_{S_1\in S_{e_1}} U^{e_1e_2}_{S_1} \tag{A.3}$$

Eq. (A.1) computes the per minute infection risk of entity e_1 between any pair of seats S_1, S_2 given the aircraft cabin configuration where pitch and width of different seat types are considered.[5] The influence of distance, individual passenger infectiousness, and mask efficiency are jointly utilized with independent distributions. The level of risk at distance zero from a contagious person of entity e_1 is denoted by parameter $\pi^{e_1}_0 \sim Beta(1,520)$. The exponentially dropping off viral infection risk with Manhattan distance d is controlled by a decay rate ω which follows the log-normal distribution $\ln(\omega) \sim N(-0.703, 0.318)$. Additionally, $\lambda \sim N(0.5, 0.1)$ represents the probability of virus transmission blockage by a seatback and depicts the virus transmission rate together with the number of rows between contagious and uninfected passengers (R). Lastly, $p_{mask} \sim N(0.7, 0.075)$ denotes the mask blocking capability.

Based on $V^{e_1}_{S_1,S_2}$, the average infection risk from contagious person at seat S_1 to other seats accessible to an infected entity e_2, whose available seat set is S_{e_2}, during a 60 min flight is computed with Eq. (A.2). Consider all the possible seats $S_1 \in S_{e_1}$ that an infector e_1 can take, the average overall risk $\beta_{e_1e_2}$ with **one** infector from entity e_1 to e_2 is obtained with Eq. (A.3). In this way, the infection risk for entity route $i-j-k$ via flight f_1, f_2 is calculated as $\beta^0_i\tau^{f_1}_i\beta_{ij}\tau^{f_2}_j\beta_{jk}$ where the expected number of infectors

[5] Airline specific seat maps are accessible through https://www.seatguru.com/airlines.

N_i at each entity i is used to derive the overall risk level (e.g. $\beta_i^0 \cdot N_i$ for entity i). Cockpit crews are assumed to be seated in the front seats and thus different infection likelihoods for the crew and passengers are computed according to B738 cabin seat map (see Fig. A.14), namely: crew-to-crew, crew-to-passenger, passenger-to-crew, and passenger-to-passenger. Due to the data limitation, the movement of cabin crews and their average contact time with passengers are not considered here.

Taking the passenger-to-crew infection as an example, assuming a passenger itinerary p (population size is 10) with initial infection probability $\beta_p^0 = 0.2$ and mask-wearing ratio 0.5 takes the same two-hour flight with a healthy pilot c, the possible seat set $\mathcal{S}_p$ is $\{3A, 3B, \dots, 30F\}$ and $\mathcal{S}_c$ is $\{1Pilot, 1Copilot\}$. Consequently, one passenger-to-crew transmission rate β_{pc} can be sampled as 0.00533. Given the parameter value $\tau_p^f = N_p \frac{FT_f}{60}^{k_1} = 10 \times \frac{120}{60}^{0.814} = 11.36$ which set $k_1 = 0.814$ as will be explained in Section 4.1.3, the infection probability of pilot c is $11.36 \times 0.00533 = 0.06$.

Appendix B. Proof of Proposition 1

Based on the formulation of Z (31)–(34), we derive its dual problem D using Lagrangian duality theory and let m, n_ω be the dual variables for Constraints (32) and (33).

$$D = \inf m\theta + \frac{1}{N}\sum_{\omega\in\Omega} n_\omega \tag{B.1}$$

$$\text{s.t.} n_\omega + m\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_\omega\|_p \geq \mathbf{v}^T\boldsymbol{\alpha} \qquad \forall\omega\in\Omega, \boldsymbol{\alpha}\in\mathcal{W} \tag{B.2}$$

$$m \in \mathbb{R}^+, n_\omega \in \mathbb{R} \tag{B.3}$$

While there exist computational difficulties in Constraints (B.2) as the cardinality is infinite, we further convert these constraints as follows:

$$n_\omega + m\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_\omega\|_p \geq \mathbf{v}^T\boldsymbol{\alpha} \qquad \forall\omega\in\Omega, \boldsymbol{\alpha}\in\mathcal{W} \tag{B.4}$$

$$\Leftrightarrow n_\omega \geq \sup_{\boldsymbol{\alpha}\in\mathcal{W}} (\mathbf{v}^T\boldsymbol{\alpha} - m\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_\omega\|_p) \qquad \forall\omega\in\Omega \tag{B.5}$$

$$\Leftrightarrow n_\omega \geq \sup_{\boldsymbol{\alpha}\in\mathcal{W}} (\mathbf{v}^T\boldsymbol{\alpha} - \max_{\|\mathbf{t}_\omega\|_{\frac{p}{p-1}}\leq m} \mathbf{t}_\omega^T(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_\omega)) \qquad \forall\omega\in\Omega \tag{B.6}$$

$$\Leftrightarrow n_\omega \geq \min_{\|\mathbf{t}_\omega\|_{\frac{p}{p-1}}\leq m} \sup_{\boldsymbol{\alpha}\in\mathcal{W}} (\mathbf{v}^T\boldsymbol{\alpha} - \mathbf{t}_\omega^T\boldsymbol{\alpha}) + \mathbf{t}_\omega^T\hat{\boldsymbol{\alpha}}_\omega \qquad \forall\omega\in\Omega \tag{B.7}$$

$$\Leftrightarrow \begin{cases} n_\omega \geq \sup_{\boldsymbol{\alpha}\in\mathcal{W}}(\mathbf{v}^T\boldsymbol{\alpha} - \mathbf{t}_\omega^T\boldsymbol{\alpha}) + \mathbf{t}_\omega^T\hat{\boldsymbol{\alpha}}_\omega, & \forall\omega\in\Omega \\ \|\mathbf{t}_\omega\|_{\frac{p}{p-1}} \leq m, & \forall\omega\in\Omega \end{cases} \tag{B.8}$$

$$\Leftrightarrow \begin{cases} n_\omega \geq \mathbf{t}_\omega^T\hat{\boldsymbol{\alpha}}_\omega - \inf \underline{\boldsymbol{\alpha}}^T\mathbf{q}_\omega, & \forall\omega\in\Omega \\ -\mathbf{q}_\omega = \mathbf{v} - \mathbf{t}_\omega, & \forall\omega\in\Omega \\ \mathbf{q}_\omega \geq \mathbf{0}, & \forall\omega\in\Omega \\ \|\mathbf{t}_\omega\|_{\frac{p}{p-1}} \leq m, & \forall\omega\in\Omega \end{cases} \tag{B.9}$$

$$\Leftrightarrow \begin{cases} n_\omega \geq \mathbf{v}^T\hat{\boldsymbol{\alpha}}_\omega - (\hat{\boldsymbol{\alpha}}_\omega - \underline{\boldsymbol{\alpha}})^T\mathbf{q}_\omega, & \forall\omega\in\Omega \\ \mathbf{q}_\omega \geq \mathbf{0}, & \forall\omega\in\Omega \\ m \geq \|\mathbf{v} + \mathbf{q}_\omega\|_{\frac{p}{p-1}}, & \forall\omega\in\Omega \end{cases} \tag{B.10}$$

The equivalence Eq. (B.6) holds due to the definition of dual norm, and Eq. (B.7) holds according to the minimax theorem. We derive the conversion from Eqs. (B.8) to (B.9) using linear duality theory for $\sup_{\boldsymbol{\alpha}\in\mathcal{W}}(\mathbf{v}^T\boldsymbol{\alpha} - \mathbf{t}_\omega^T\boldsymbol{\alpha})$ on $\mathcal{W} = \{\boldsymbol{\alpha} \geq \underline{\boldsymbol{\alpha}}\}$.

To optimize Eq. (B.1), we observe that both m and n_ω increase monotonically with $\mathbf{q}_\omega$, then we can set $\mathbf{q}_\omega = 0$ to get the optimal condition as follows:

$$\begin{cases} n_\omega \geq \mathbf{v}^T\hat{\boldsymbol{\alpha}}_\omega, & \forall\omega\in\Omega \\ m \geq \|\mathbf{v}\|_{\frac{p}{p-1}}, & \forall\omega\in\Omega \end{cases} \tag{B.11}$$

Therefore, according to the definition of polynomial norm, we can derive a closed-form solution as $D^* = \theta|r|^{\frac{p-1}{p}} + \frac{1}{N}\sum_{\omega\in\Omega}\mathbf{v}^T\hat{\boldsymbol{\alpha}}_\omega$. Due to the strong duality theorem, the optimal solution of the original problem Z^* is the same as D^* and we complete the proof.

References

Abdelghany, A., Ekollu, G., Narasimhan, R., Abdelghany, K., 2004. A proactive crew recovery decision support tool for commercial airlines during irregular operations. Ann. Oper. Res. 127 (1), 309–331.

Arıkan, U., Gürel, S., Aktürk, M.S., 2017. Flight network-based approach for integrated airline recovery with cruise speed control. Transp. Sci. 51 (4), 1259–1287.

Barnett, A., Fleming, K., 2022. Covid-19 infection risk on US domestic airlines. Health Care Manag. Sci. 1–16.

Barnhart, C., Fearing, D., Vaze, V., 2014. Modeling passenger travel and delays in the national air transportation system. Oper. Res. 62 (3), 580–601.
Birolini, S., Jacquillat, A., Cattaneo, M., Antunes, A.P., 2021. Airline Network Planning: Mixed-integer non-convex optimization with demand–supply interactions. Transp. Res. B 154, 100–124.
Bratu, S., Barnhart, C., 2006. Flight operations recovery: New approaches considering passenger recovery. J. Sched. 9 (3), 279–298.
Chu, Y., Xia, Q., 2004. Generating benders cuts for a general class of integer programming problems. In: International Conference on Integration of Artificial Intelligence (AI) and Operations Research (OR) Techniques in Constraint Programming. Springer, pp. 127–141.
Cristian Ionut Panait, M.E.F.., 2020. Guidance on the Management of Crew Members in Relation to the COVID-19 Pandemic. Technical Report, European Union Aviation Safety Agency.
Ding, Y., Wandelt, S., Sun, X., 2021. TLQP: Early-stage transportation lock-down and quarantine problem. Transp. Res. C 129, 103218.
Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20 (5), 533–534.
Eggenberg, N., Salani, M., Bierlaire, M., 2010. Constraint-specific recovery network for solving airline recovery problems. Comput. Oper. Res. 37 (6), 1014–1026.
European Commission, 2020. Interpretative Guidelines on EU Passenger Rights Regulations in the Context of the Developing Situation with Covid-19. The European Commission, URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52020XC0318(04).
Evler, J., Asadi, E., Preis, H., Fricke, H., 2021. Airline ground operations: Schedule recovery optimization approach with constrained resources. Transp. Res. C 128, 103129.
Freedman, D.O., Wilder-Smith, A., 2020. In-flight transmission of SARS-CoV-2: a review of the attack rates and available data on the efficacy of face masks. J. Travel Med. 27 (8), taaa178.
Garrow, L.A., Lurkin, V., Marla, L., 2022. Airline OR innovations soar during COVID-19 recovery. Oper. Res. Forum 3 (1).
Gounaris, C.E., Wiesemann, W., Floudas, C.A., 2013. The robust capacitated vehicle routing problem under demand uncertainty. Oper. Res. 61 (3), 677–693.
Government of Canada, 2021. Evidence on the risk of COVID-19 transmission in flight: update 3. URL: https://www.canada.ca/content/dam/phac-aspc/documents/services/diseases/2019-novel-coronavirus-infection/canadas-reponse/summaries-recent-evidence/evidence-risk-covid-19-transmission-flight-update-3/evidence-risk-covid-19-transmission-flight-update-3.pdf.
Hertzberg, V.S., Weiss, H., 2016. On the 2-row rule for infectious disease transmission on aircraft. Ann. Global Health 82 (5), 819–823.
Hertzberg, V.S., Weiss, H., Elon, L., Si, W., Norris, S.L., Team, F.R., 2018. Behaviors, movements, and transmission of droplet-mediated respiratory diseases during transcontinental airline flights. Proc. Natl. Acad. Sci. 115 (14), 3623–3627.
Hu, Y., Zhang, P., Fan, B., Zhang, S., Song, J., 2021. Integrated recovery of aircraft and passengers with passengers' willingness under various itinerary disruption situations. Comput. Ind. Eng. 161, 107664.
Huang, Z., Luo, X., Jin, X., Karichery, S., 2021. An iterative cost-driven copy generation approach for aircraft recovery problem. European J. Oper. Res..
Ito, K., Piantham, C., Nishiura, H., 2022. Relative instantaneous reproduction number of Omicron SARS-CoV-2 variant with respect to the Delta variant in Denmark. J. Med. Virol. 94 (5), 2265–2268.
Kelly, D., Bambury, N., Boland, M., 2021. In-flight transmission of wild-type SARS-CoV-2 and the outbreak potential of imported clusters of COVID-19: a review of published evidence. Glob. Health 17 (1), 1–5.
Lau, M.S., Grenfell, B., Thomas, M., Bryan, M., Nelson, K., Lopman, B., 2020. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. Proc. Natl. Acad. Sci. 117 (36), 22430–22435.
Lee, J., Marla, L., Jacquillat, A., 2020. Dynamic disruption management in airline networks under airport operating uncertainty. Transp. Sci. 54 (4), 973–997.
Lettovskỳ, L., Johnson, E.L., Nemhauser, G.L., 2000. Airline crew recovery. Transp. Sci. 34 (4), 337–348.
Liang, Z., Xiao, F., Qian, X., Zhou, L., Jin, X., Lu, X., Karichery, S., 2018. A column generation-based heuristic for aircraft recovery problem with airport capacity constraints and maintenance flexibility. Transp. Res. B 113, 70–90.
Luttmann, A., 2019. Are passengers compensated for incurring an airport layover? Estimating the value of layover time in the US airline industry. Econ. Transp. 17, 1–13.
Maher, S.J., 2015. A novel passenger recovery approach for the integrated airline recovery problem. Comput. Oper. Res. 57, 123–137.
Maher, S.J., 2016. Solving the integrated airline recovery problem using column-and-row generation. Transp. Sci. 50 (1), 216–239.
Mancini, S., 2016. A real-life multi depot multi period vehicle routing problem with a heterogeneous fleet: Formulation and adaptive large neighborhood search based matheuristic. Transp. Res. C 70, 100–112.
Mangili, A., Gendreau, M.A., 2005. Transmission of infectious diseases during commercial air travel. Lancet 365 (9463), 989–996.
Marla, L., Vaaben, B., Barnhart, C., 2017. Integrated disruption management and flight planning to trade off delays and fuel burn. Transp. Sci. 51 (1), 88–111.
Namilae, S., Derjany, P., Mubayi, A., Scotch, M., Srinivasan, A., 2017. Multiscale model for pedestrian and infection dynamics during air travel. Phys. Rev. E 95 (5), 052320.
Nissen, R., Haase, K., 2006. Duty-period-based network model for crew rescheduling in European airlines. J. Sched. 9 (3), 255–278.
Petersen, J.D., Sölveling, G., Clarke, J.-P., Johnson, E.L., Shebalov, S., 2012. An optimization approach to airline integrated recovery. Transp. Sci. 46 (4), 482–500.
Poletto, C., Gomes, M., y Piontti, A.P., Rossi, L., Bioglio, L., Chao, D.L., Longini, Jr., I., Halloran, M.E., Colizza, V., Vespignani, A., 2014. Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. Eurosurveillance 19 (42), 20936.
Ribeiro, N.A., Jacquillat, A., Antunes, A.P., 2019. A large-scale neighborhood search approach to airport slot allocation. Transp. Sci. 53 (6), 1772–1797.
Saddoune, M., Desaulniers, G., Soumis, F., 2013. Aircrew pairings with possible repetitions of the same flight number. Comput. Oper. Res. 40 (3), 805–814.
Salari, M., Milne, R.J., Delcea, C., Kattan, L., Cotfas, L.-A., 2020. Social distancing in airplane seat assignments. J. Air Transp. Manag. 89, 101915.
Schultz, M., Evler, J., Asadi, E., Preis, H., Fricke, H., Wu, C.-L., 2020. Future aircraft turnaround operations considering post-pandemic requirements. J. Air Transp. Manag. 89, 101886.
Schultz, M., Soolaki, M., 2021. Analytical approach to solve the problem of aircraft passenger boarding during the coronavirus pandemic. Transp. Res. C 124, 102931.
Shafipour-Omrani, B., Komijan, A.R., Sadjadi, S.J., Khalili-Damghani, K., Ghezavati, V., 2021. A flexible mathematical model for crew pairing optimization to generate n-day pairings considering the risk of COVID-19: a real case study. Kybernetes 51 (12), 3545–3573.
Shebalov, S., Klabjan, D., 2006. Robust airline crew pairing: Move-up crews. Transp. Sci. 40 (3), 300–312.
Sinclair, K., Cordeau, J.-F., Laporte, G., 2014. Improvements to a large neighborhood search heuristic for an integrated aircraft and passenger recovery problem. European J. Oper. Res. 233 (1), 234–245.
Su, Y., Xie, K., Wang, H., Liang, Z., Chaovalitwongse, W.A., Pardalos, P.M., 2021. Airline disruption management: A review of models and solution methods. Engineering 7 (4), 435–447.
Suau-Sanchez, P., Voltes-Dorta, A., Cuguero-Escofet, N., 2020. An early assessment of the impact of COVID-19 on air transport: Just another crisis or the end of aviation as we know it? J. Transp. Geogr. 86, 102749.
Subramanyam, A., Repoussis, P.P., Gounaris, C.E., 2020. Robust optimization of a broad class of heterogeneous vehicle routing problems under demand uncertainty. INFORMS J. Comput. 32 (3), 661–681.
Sun, X., Wandelt, S., Zhang, A., 2020. How did COVID-19 impact air transportation? A first peek through the lens of complex networks. J. Air Transp. Manag. 89, 101928.
Sun, X., Wandelt, S., Zhang, A., 2021a. Delayed reaction towards emerging COVID-19 variants of concern: Does history repeat itself? Transp. Res. A: Policy Pract. 152, 203–215.

Sun, X., Wandelt, S., Zhang, A., 2021b. On the degree of synchronization between air transport connectivity and COVID-19 cases at worldwide level. Transp. Policy 105, 115–123. http://dx.doi.org/10.1016/j.tranpol.2021.03.005, URL: https://www.sciencedirect.com/science/article/pii/S0967070X21000652.

Sun, X., Wandelt, S., Zhang, A., 2022a. COVID-19 pandemic and air transportation: Summary of recent research, policy consideration and future research directions. Transp. Res. Interdiscip. Perspect. 16, 100718.

Sun, X., Wandelt, S., Zhang, A., 2022b. STARTUPS: Founding airlines during COVID-19-A hopeless endeavor or an ample opportunity for a better aviation system? Transp. Policy 118, 10–19.

Sun, X., Wandelt, S., Zheng, C., Zhang, A., 2021c. COVID-19 pandemic and air transportation: Successfully navigating the paper hurricane. J. Air Transp. Manag. 102062.

Thengvall, B.G., Bard, J.F., Yu, G., 2003. A bundle algorithm approach for the aircraft schedule recovery problem during hub closures. Transp. Sci. 37 (4), 392–407.

Vink, J., Santos, B.F., Verhagen, W.J., Medeiros, I., et al., 2020. Dynamic aircraft recovery problem-an operational decision support framework. Comput. Oper. Res. 117, 104892.

Xu, Y., Wandelt, S., Sun, X., 2021. Airline integrated robust scheduling with a variable neighborhood search based heuristic. Transp. Res. B 149, 181–203.

Zhang, D., Lau, H.H., Yu, C., 2015. A two stage heuristic algorithm for the integrated aircraft and crew schedule recovery problems. Comput. Ind. Eng. 87, 436–453.

Zhang, Y., Zhang, Z., Lim, A., Sim, M., 2021. Robust data-driven vehicle routing with time windows. Oper. Res. 69 (2), 469–485.