Information Fusion 101 (2024) 101983

Contents lists available at ScienceDirect

Information Fusion

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

Fuel consumption prediction for pre-departure flights using attention-based multi-modal fusion

Yi Lin [a], Dongyue Guo [a], Yuankai Wu [b], Lishuai Li [c,d], Edmond Q. Wu [e], Wenyi Ge [f,*]

[a] College of Computer Science, Sichuan University, Chengdu, 610065, China
[b] Department of Civil and Environmental Engineering, McGill University, Montreal, H3A 0C3, Canada
[c] Aerospace Engineering, Delft University of Technology, The Netherlands
[d] School of Data Science, City University of Hong Kong, Hong Kong SAR
[e] Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai Jiao Tong University, Shanghai, 200240, China
[f] School of Computer Science, Chengdu University of Information Technology, Chengdu, 610225, China

ARTICLE INFO

Keywords:
Attention mechanism
ConvLSTM
Graph convolutional network
Multi-modal fusion
Pre-departure fuel consumption prediction

ABSTRACT

Improper fuel loading decision results in carrying excessive dead weight during flight operation, which will burden the airline operation cost and cause extra waste emission. Existing works mainly focused on the post-event fuel consumption based on flight trajectory. In this work, a novel deep learning model, called FCPNet, is proposed to achieve the fuel consumption prediction (FCP) before the flight departure. Considering the influential factors for aircraft performance, the multi-modal information sources, including the planned route, weather information, and operation details, are selected as the model input to predict fuel consumption. Correspondingly, three modules are innovatively proposed to learn embedding features from multi-modal inputs. Based on the planned route, the graph convolutional network is proposed to mine the spatial correlations in the non-Eulerian route network. Considering the grid attributes of the weather information, the ConvLSTM is applied to learn abstract representations from both the temporal and spatial dimensions, in which the three-dimensional convolution neural networks are also designed to fine-tune intermediate feature maps. The fully connected layer is also proposed to learn informative features from operation details. Finally, an attention-based fusion network is presented to generate the final embedding by considering the unique contributions of the multi-modality sources, which are further applied to predict flight fuel consumption. A binary encoding representation is proposed to formulate the FCP task as a multi-binary classification problem. The proposed model is validated on a real-world dataset, and the results demonstrate that it outperforms other baselines, i.e., achieving a 6.50% mean absolute percentage error, which can practically support the airline operation and global emission control before flight departure.

ACRONYMS

AttFN	attention-based fusion network
BE	binary encoding
ConvLSTM	convolutional LSTM Network
FC	fully connected layer
FCP	fuel consumption prediction
FRE	flight route embedding
GCN	graph convolutional network
LSTM	long short-term memory
mBC	multi-binary classification
OD	origin-destination
ODE	operation detail embedding
PN	prediction network
WIE	weather information embedding

1. Introduction

The carbon-neutral is of great importance to the target of global temperature control. Currently, transportation-related industries account for over 30% of the global energy demand and contribute more than 20% of the global carbon emission [1]. To be specific, about 11.6% of the transportation emissions (2% of global) are generated by the operation of the aviation industry, i.e., the top among different

* Corresponding author.
E-mail address: gewenyi15@cuit.edu.cn (W. Ge).

https://doi.org/10.1016/j.inffus.2023.101983
Received 27 July 2023; Received in revised form 16 August 2023; Accepted 21 August 2023
Available online 22 August 2023

transportation modes [2]. With the booming of the global economy, the number of passengers for flight traveling will double to 8.2 billion in 2037 according to the investigation by the International Air Transport Association [3]. Thus, it can be seen that optimizing aviation emissions is of great potential to achieve the carbon-neutral target, which is also able to greatly reduce airline costs.

For a certain flight, the final total fuel loading is planned by various categories, including mission fuel, alternate fuel, reserve fuel, contingency fuel, etc. In general, the International Civil Aviation Organization published a series of aviation regulations to support the planning for alternate fuel, reserve fuel, and contingency fuel [4]. For instance, the alternate fuel must at least meet the requirement of the 45-minute flight cruise for safety issues, typically for the landing at an alternative airport. As to the mission fuel, the airline operation empirically estimates the fuel consumption based on the flight route (total journey and flight altitude) before the flight departure. As investigated in [5], due to improper fuel loading decisions before flight departure, about 4.5% of the total fuel is wasted by carrying "dead weight" of the unused fuel, which causes an extra $230 million expense and corresponding carbon emissions.

Considering the aforementioned issues, it is highly required to provide an accurate flight fuel consumption prediction (FCP) tool that can be applied to determine the fuel loading (mainly for mission fuel) before the flight departure. The overall goal of the flight fuel consumption prediction is to reduce the fuel loading while ensuring flight safety. Existing FCP approaches mainly focused on aircraft or engine designs [6], flight operations [7], and trajectory simulation [8], in which the data-driven mechanism was applied to improve the prediction performance. However, almost all those approaches depended on trajectory details of the flight operation as a post-analysis tool, which is usually intractable and hard to accurately obtain before the flight departure, and the prediction models failed to be migrated to a real application. Therefore, in this work, we attempt to achieve the FCP task before the flight departure to support the fuel loading decisions and hence reduce the unnecessary dead weight, which further benefits to saving the operational cost of airlines and global waste emissions.

To this end, a novel approach is proposed to achieve the FCP task mainly based on the planned flight route and other multi-source operational details, called FCPNet. Meanwhile, considering that the weather information is a decisive factor in aircraft performance for the fuel consumption during the flight cruise [9], the wind velocity is selected as an auxiliary input to improve the final performance. In general, the flight route for a certain origin-destination (OD) airport pair usually shares the same patterns for most operations, but the total amounts of fuel consumption for them are extremely different. This result indicates that, except for the flight route, the final fuel consumption is also influenced by other required information, which is the reason why we consider the multi-modality information to enhance the prediction performance. Inspired by popular applications of deep learning techniques in the aviation industry [10], in this work, the whole FCP model is constructed by neural architectures.

Considering the different modalities of the flight route, weather information, and other operation details, three independent neural modules are designed to learn informative patterns from the raw input information. The planned route of the flight is firstly illustrated by a route graph, which is further fed into a graph convolutional network (GCN)-based architecture to generate an embedding vector. Regarding the weather information, the wind velocity is depicted by a 3-dimensional (3D) grid to characterize the influence on the fuel consumption, which is also fed into the ConvLSTM-based network to generate an embedding by combining it with the temporal dimension. For operational details, the fully connected (FC) layer-based network is constructed to learn informative embeddings to support the prediction task. Finally, a fusion module is designed to consider the task-oriented features to achieve the final FCP task, in which the attention mechanism is proposed to learn desired fusion weights by data-driven optimization.

To address the high dimension of the input information, a binary encoding (BE) representation is proposed to convert the integer-based fuel consumption to a binary code (with only 0 and 1). The primary purpose is to formulate a smooth model architecture, from a 1-dimensional scalar to a multi-dimensional binary vector. With the proposed BE representation, in this work, the FCP task is re-defined as multi-binary classification (mBC) tasks, in which both the classification and regression loss are integrally considered to optimize the proposed model.

A real-world dataset is built to validate the proposed approach and the final results demonstrate that the proposed model is able to achieve the flight fuel consumption prediction with considerably higher performance without extra safety issues, i.e., 6.50% MAPE (mean absolute percentage error). All in all, the contribution of this work is summarized as follows:

1) A deep learning-based model, called FCPNet, is proposed to predict fuel consumption to support the fuel loading decisions before flight departure. The proposed model can achieve comparable performance only based on the planned route information, and the performance can be significantly enhanced by considering the multi-modality inputs.
2) Considering the multi-modalities of the input data, the GCN- and ConvLSTM-based networks are designed to learn informative patterns from flight route sequence and weather information, which further supports the final prediction task.
3) An attention-based fusion module is proposed to learn task-oriented features for the fuel consumption prediction task, which clarifies the contributions of the multi-modality data from both the modality and embedding dimension by different flight operations.
4) A binary encoding representation is proposed to prevent the sharp output dimension reduction of the regression task, which further enhances the model convergence and training ability. The FCP task is finally achieved by an mBC-related task with both classification and regression loss.
5) Extensive experiments are conducted to validate the proposed approach on a real-world dataset, and the underlying patterns of different modules are also confirmed by experiments.

The rest of this work is organized as follows. The related works about the research topic are reviewed in Section 2. The details of the proposed framework are provided in Section 3. In Section 4, we list the experimental configurations for validating the proposed approach. The experimental results are reported and discussed in Section 5. Finally, this paper is concluded in Section 6.

2. Related works

2.1. Fuel consumption prediction

With the booming of the civil aviation industry, the fuel consumption prediction task is attracting increasing research attention from all over the world, but it is still an emerging field for academic research. The early-stage works were reviewed in [11], concerning data, method and research topics. The fuel consumption for the flight in the uplift phase was explored to achieve the target of landing with empty waste fuel for US civil aviation [12]. The flight gate-in fuel consumption was studied to measure the predictability of the civil aviation system, which further considered the operational cost reduction for several airlines [5]. The fuel consumption prediction was also studied for low-cost airlines considering the flight delay [13]. Based on the aircraft performance, the artificial neural network was applied to estimate the fuel consumption for a certain aircraft [14]. Similarly, the fuel flow dynamics were introduced to guide the modeling of fuel consumption using a data-driven mechanism [6]. Specifically, the data-driven model was also proposed to predict the fuel burn for the climb and approach phases

considering aircraft engines [15]. As a post-analysis, the flight operation quality assurance (FOQA) data was used to estimate the fuel consumption as a data fitting problem [16], a similar research in [17] based on onboard flight data recorder (FDR). A neural network-based approach was proposed to predict flight delay and fuel consumption [4]. The trajectory simulation technique was applied to predict the fuel estimation by considering the OD airport pairs and aircraft type [8], and the final performance is impacted by the accuracy of the trajectory prediction. The ensemble learning was applied to achieve the fuel burn prediction and uncertainty estimation, which is further able to improve fuel efficiency and also reduce operation cost [18]. Except for air transport, the FCP task is a hot research topic in ground transport, including fuel evaluation for mixed traffic [19] and energy-oriented lane-change strategy [20], which can provide successful cases to this study.

In summary, existing related works achieve the fuel consumption task based on the aircraft performance and post-flight trajectory, serving as the post-analysis tool for airline operations. However, to effectively support the decision of the fuel loading for airlines, it is highly required to determine the fuel consumption before the flight departure. Existing tools failed to achieve a desired performance since the flight trajectory is usually intractable and hard to accurately obtain before the flight departure, which further inspires us to reconsider the FCP task for pre-departure implementation.

2.2. Deep learning-related techniques

In this section, the related techniques of deep learning approaches for transportation research are reviewed. Deep learning-based models were widely studied to address the ground and aviation transportation issues [21,22]. The convolution neural network (CNN) was proposed to achieve the ground traffic flow prediction [23]. Meanwhile, considering the temporal transition patterns of the traffic data, the recurrent neural network (RNN) was applied to predict the traffic characteristic from the perspective of temporal modeling, including gate recurrent unit and long short-term memory [24]. Lately, inspired by the nature of the road network, the GCN-based network was proposed to mine abstract representations in a non-Eulerian space [25], supporting the final traffic prediction task. Similarly, the temporal modeling was also integrated into GCN models to consider both the temporal and spatial correlations, and hence generates the temporal GCN model (TGCN) [26]. The CNN and LSTM blocks were also combined to learn temporal and spatial patterns to achieve the traffic prediction task in an integrated manner [10].

In general, the multi-source data was utilized to achieve the final prediction task in transportation research fields, such as the weather or prior rules [27], etc. The multi-modal fusion strategy was studied to enhance the detection performance in surveillance scenes using deep learning techniques [28]. As to speech recognition, visual information was usually selected as the auxiliary feature to improve the model performance [29]. The attention mechanism was proposed to learn task-specific weights for different inputs or components. The attention modules were firstly designed to achieve natural language processing tasks, such as the Transformer [30]. The squeeze-and-excitation network [31], i.e., SENet, was designed to guide the model to focus on desired feature channels to improve the computer vision task. The convolutional block attention module [32] was designed to capture the channel and spatial attentions, which can be integrated into various neural architectures to obtain performance improvements. The attention mechanism was also applied to enhance the performance of the traffic prediction tasks [33]. A dynamic multi-modal fusion strategy was studied to adaptively consider multi-modal features across several tasks [34].

3. Methodology

3.1. The proposed framework

In this work, we attempt to achieve the FCP task to support the fuel loading decisions for airline operations before the flight departure. In general, the available information for a pre-departure flight comprises the planned route, weather information, and required operation details. Considering the data representation of the three kinds of information, the multi-path architecture is designed to process the multi-modality information, in which corresponding neural networks are applied to extract informative features to support the prediction task. In this work, the whole model is implemented by deep neural networks, including a GCN-based flight route embedding (FRE) network, ConvLSTM-based weather information embedding (WIE) network, FC-based operation detail embedding (ODE) network, and an attention-based fusion network (AttFN), and an FC-based prediction network (PN). The whole architecture of the proposed model is sketched in Fig. 1, and each module can be briefly described as follows:

a) FRE module: the primary purpose of this module is to learn informative features from the planned route of a given flight. Considering the graph nature of the route network, the GCN block serves as the fundamental block, which is able to cope with the pattern specificities of a non-Eulerian space. The node-level features are extracted from GCNs to finally generate an embedding vector.
b) WIE module: this module aims to mine the significant features from the temporal weather information on certain airspaces and flight routes. Considering the input representation, the ConvLSTM block is applied to generate compressed feature maps from the 3D grid data. The 3D convolutional operation (3D-Conv) is also introduced to compress the data dimension by the data-driven mechanism. The final output is expected to be a compact embedding to represent the overall temporal and spatial correlations of weather information on fuel consumption.
c) ODE module: this module focuses on extracting influential patterns from operation details denoted by multi-dimensional indicators. As the operation details of a flight involve different kinds of information, in the module, the FC layers are densely appended to generate a modality-related embedding.
d) AttFN module: considering the specificities of different input modalities, the independent embeddings generated from three modules may not be optimal for the final task. To extract high-level task-oriented representations, a feature fusion network is innovatively designed to refine the generated embeddings, in which the attention mechanism is applied to learn weights for both modality and embedding correlations.
e) PN module: the attentive embeddings generated by the aforementioned four blocks are further fed into the FC-based prediction network to estimate the fuel consumption for a given flight.

In general, the planned flight route for a certain O-D pair usually shares the same patterns for most operations, but the total amounts of fuel consumption for them are extremely different. This indicates that, except for the flight route, the fuel consumption is also influenced by other required information, which is the reason why we consider the multi-modal information to enhance the prediction performance. In summary, the proposed model takes both the static (FRE and ODE) and dynamic (WIE) factors to achieve the FCP task. The neural architecture is designed to process the corresponding data organization in this work, i.e., GNN for route network, ConvLSTM for sequential weather images and FC for discrete vectors. Meanwhile, to learn discriminative patterns on input samples, an attention-based multi-modal fusion is innovatively designed to extract fine-grained features to support the FCK task. Mathematically, the inference rules of the proposed model can be briefly described as (1–3), described as an embedding, fusion and prediction

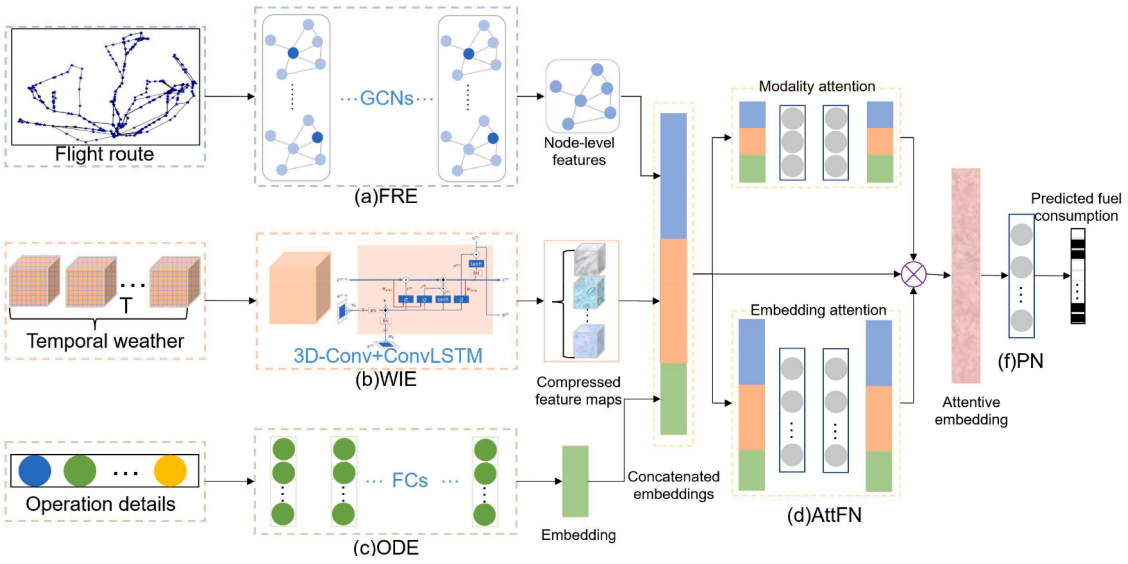

Fig. 1. The model architecture of the proposed FCPNet.

procedure. The E_{fre}, E_{wie} and E_{ode} denotes the intermediate embeddings generated by corresponding modules, respectively. The *AttE* is the attentive embeddings with attention mechanism, and the final FCP is achieved by the PN network.

$$E_{fre} = FRE(X_r), E_{wie} = WIE(X_w), E_{ode} = ODE(X_o) \tag{1}$$

$$AttE = AttFN\left(E_{fre}, E_{wie}, E_{ode}\right) \tag{2}$$

$$FCP = PN(AttE) \tag{3}$$

3.2. Model inputs

Before the flight departure, the available information consists of the planned flight route, predicted weather information, and operation details. The model inputs are described as follows:

3.2.1. The flight route

For the flight operation, the flight management system predicts its trajectory profiles based on the planned route. In general, the details consist of the flyover positions, altitude, speed, time, travel distance (based on departure airport), and flying status, as shown below:

$$F = [lon, lat, alt, speed, time, dis, sta]^T \tag{4}$$

The flight route determines the global trajectory of the flight operation, including the flight distance, flight time, flight profile (altitude), etc., which are primary trajectory-related contributors to fuel consumption. By referring to the ground transportation system, the planned flight route is illustrated by a graph (called route graph), in which the waypoints and corresponding connecting routes are regarded as the node and edge of the graph, respectively.

A route graph is denoted by $G = (V,E,A)$. The V is a finite set with N nodes, i.e., waypoints in this work. E saves the edges (route segments) connecting different nodes, and A denotes the adjacent matrix to illustrate the graph attributes. The element of the adjacent matrix $a_{ij} \in A$ is defined as below, which $i,j \in V$ denotes the graph node, and d_{ij} is the distance between the node i and j.

$$a_{ij} = \begin{cases} \exp(-d_{ij}) & e_{ij} \in E \\ 0 & e_{ij} \notin E \end{cases} \tag{5}$$

Considering the flight route, the proposed model can be regarded as an OD-wise implementation, in which the attributes of the flight route are dependent on the waypoint-level information. Based on the flight operation, some typically planned routes are designed for a certain OD pair, which provides a solid foundation for building a route graph and further supports the training for the proposed model. Finally, the input of the flight route modality is with the shape of $(N, |F|)$, where $|F|$ is the feature dimension.

3.2.2. The weather information

The weather information is critical to the real-time flight operation, concerning route planning, flight profile, and the impacts on aircraft performance. Generally, in the mentioned factors, route planning and profile optimization are regarded as the pre-disposition strategy, which is ultimately reflected in the change of the flight path, and further relates to the aircraft performance. Given the planned route before the flight departure, the fuel consumption of the flight operation is determined by the fuel consumption per distance (miles or kilometers) based on aircraft performance caused by the influence of the weather information, typically the wind velocity.

In this work, certain airspace is firstly illustrated as a 3D grid, as shown in Fig. 2. The three dimensions denote the longitude, latitude, and altitude, respectively. For a certain time-step, the wind velocity is further represented as a vector for each grid:

$$v = \left(v_x^t, v_y^t\right) \tag{6}$$

where t is the time instant. The v_x^t and v_y^t are the velocity projection on the longitude and latitude dimensions, respectively. A positive v_x^t indicates that the wind is from west to east, and a negative for east to west.

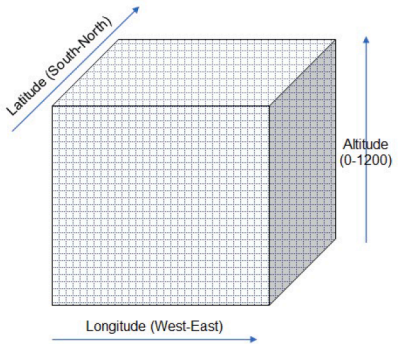

Fig. 2. The representation of the wind velocity.

The v_y follows similar rules. It can be seen that the wind velocity is a kind of temporal and spatial related pattern on the flight route, which provides influential factors to the FCP task for certain airspaces or routes at a given time instant.

Finally, considering the temporal characteristics of the flight operation, the input shape of the weather information in this work is represented by:

$$Wea = (T, Z, X, Y, 2) \tag{7}$$

where T is the time span corresponding to the flight operations. Z, X and Y are the shape of the altitude, longitude, and latitude of the airspace grid, respectively. Referring to Eq. (6), the number of dimensions for the wind velocity v is 2 in this work.

3.2.3. The operation details

As a pre-departure prediction task, the flight-level related information, i.e., operational details, should also be considered to estimate the fuel consumption. The operational details include general information about the flight operation, such as the aircraft type, the departure and destination airports, the flight date, etc., which are primary plan-related contributors to fuel consumption.

As listed in Table 1, each indicator is represented by a one-hot vector, i.e., the operation details are concatenated as a 54-dimensional one-hot vector, with the option of only 0 or 1. Details can be found below:

a) AT is the aircraft type, denoting the overall aircraft performance with respect to different real-time factors during the flight operation, such as A320, B747, etc. Currently, a total of 13 aircraft types are in our dataset.
b) ADEP and ADES concern the airport class for the departure and destination, respectively. The airport class is divided into four categories in China. In general, the airport class indicates the traffic density, which further relates to the flight profile and flight time, etc.
c) DoW denotes the day of the week, i.e., 7-dimension. Since the flight schedule is weekly executed in China, the DoW is expected to represent global traffic flow patterns for certain OD pairs by considering multi-modality sources.
d) Similar to the DoW, the clock label implies the global traffic flow in a day, i.e., peak hours.
e) The holiday flag shows whether the flight date is a holiday or not, which further impacts the traffic density. In general, the traffic density during the holiday is much higher than that on the other days.

3.3. Model architecture

3.3.1. The FRE module

As illustrated before, the input of the FRE module is a route graph that represents the planned route of a flight in a non-Eulerian space. Inspired by this fact, the graph neural network (GNN) is naturally selected as the basic block in this module. Meanwhile, since the spatial correlations of the waypoint position in the 3D earth space play a significant role in depicting the flight path, the convolutional operation is also integrated into the GNN block to formulate the GCN block, as listed below.

$$\begin{aligned} F_O &= \delta\left(\widetilde{D}^{-0.5}\widetilde{A}\widetilde{D}^{-0.5}F_I\Theta\right) \\ \widetilde{A} &= A + I_N, \widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij} \end{aligned} \tag{8}$$

where $F_I \in \Re^{N\times C}$ and $F_O \in \Re^{N\times F}$ are the input and output feature, respectively, and here C and F denote the feature dimension and filter number. $\Theta \in \Re^{C\times F}$ is the learnable weights. $\widetilde{D}$ is the graph degree matrix with the addition of self-loops. $\widetilde{A}$ is the is the graph adjacency matrix with the addition of self-loops, which can be obtained based on the adjacent matrix A and identity matrix I_N. Specifically, the elements in $\widetilde{D}$ can be calculated using the sum operation on the corresponding row in $\widetilde{A}$.

As shown in Fig. 1(a), the FRE module is constructed by stacking GCN layers, in which the PReLU serves as the activation function. The node-level features are generated by GCN layers to imply the graph representations, which is finally converted into the embedding vector to support the fusion operation with other information modality.

3.3.2. The WIE module

Considering that the input is the grid-based data with both the temporal and 3D spatial dimensions, the ConvLSTM block is applied to construct the WIE module in this work. Learning from other related works [10], the ConvLSTM block has the ability to mine salient transition patterns from both the temporal and spatial dimensions in an integral manner, which is particularly important to process temporal and spatial data, such as the weather information. In this work, the ConvLSTM block is able to convert the large-scale wind velocity into a compatible high-level representation, which denotes the influence of weather information on aircraft performance to support the prediction task. The inference rules of the ConvLSTM block are shown below:

$$\begin{aligned} I^t &= f\left(W_{ix} * x^t + W_{ih} * h^{t-1} + W_{ic} \odot C^{t-1} + b_i\right) \\ F^t &= f\left(W_{fx} * x^t + W_{fh} * h^{t-1} + W_{fc} \odot C^{t-1} + b_f\right) \end{aligned} \tag{9}$$

$$C^t = F^t \odot C^{t-1} + I^t \odot g\left(W_{cx} * x^t + W_{ch} * h^{t-1} + b_c\right) \tag{10}$$

$$O^t = f\left(W_{ox} * x^t + W_{oh} * h^{t-1} + W_{oc} \odot C^t + b_o\right) \tag{11}$$

$$h^t = O^t \odot g(C^t) \tag{12}$$

where the I^t, F^t, C^t, O^t, and h^t are the hidden features of the input gate, forget gate, cell, output gate, and the hidden unit, respectively. t is the time instant. $\boldsymbol{x}^t$ is the input feature, and $\boldsymbol{b}_\cdot$ is the learnable bias for different blocks. $\boldsymbol{W}_\cdot$ is the learnable weight between two certain blocks, e.g., W_{ic} is the weight matrix for the information flow from the input gate and the cell. f and g denote activation functions. $*$ and $\odot$ are the convolutional and element-wise product operations, respectively.

In addition, to fine-tune the extracted features, the 3D convolutional operation is followed by each ConvLSTM block, which aims to retain informative patterns and compress the data dimension to enhance the model training. Finally, the extracted feature maps are squeezed into an embedding vector that is expected to represent the impacts of the wind velocity on the flight route, which have a salient influence on the FCP task.

3.3.3. The AttFN module

In this work, the FCP task is achieved by capturing the underlying patterns from different kinds of input information with certain modalities. Three independent modules are designed to learn informative features from different modalities, in which the embedding vectors are generated to support the final prediction task. It is well received that the unique contributions in a multi-modality system should be considered to understand the data patterns for enhancing the final performance. To this end, a fusion network is designed to recalibrate the embeddings

Table 1
Indicators for operation details.

Indicators	Description	#Dimension
AT	The representations of overall performance based on the aircraft type	13
ADEP	The class for the departure airport	4
ADES	The class for the destination airport	4
DoW	Day of week	7
Clock	Hour of day	24
Holiday	Holiday flag	2

from different inputs, which finally generates an attentive embedding for the subsequent module. Considering the diversities of the multi-modality data, the attention mechanism is proposed to optimize the feature weights in a learnable manner. Facing the representations of the multi-modality inputs, in this work, two parallel paths are dedicatedly developed to generate an optimal embedding to support the final task, as shown below:

a) modality attention: this path is to consider certain contributions of different inputs information by learning modality attention weights, i.e., inter-modality. This path is expected to capture the importance of different inputs (modality) on the condition of different data patterns and distributions, and an optimal convergence is expected to be reached by the data-driven mechanism.
b) embedding attention: this path is to consider contributions of feature dimensions for each modality by learning embedding attention weights, i.e., intra-modality. This path is mainly to ensure the robustness of the calibrated feature embedding, which prevents the representation collapse caused by any local feature perturbation.
c) feature embedding: finally, the attentive embedding is generated by combining the input features, modality attention weights, and embedding attention weights using the operation of the element product.

Given the inputs E_{fre}, E_{wie} and E_{ode}, the inference rules of the attention-based fusion are in (13), and the final embedding is generated by (14) and (15):

$$\begin{bmatrix}\alpha_{fre},\alpha_{wie},\alpha_{ode}\end{bmatrix} = MA\left(E_{fre},E_{wie},E_{ode}\right)$$
$$\begin{bmatrix}\beta_{fre},\beta_{wie},\beta_{ode}\end{bmatrix} = EA\left(E_{fre},E_{wie},E_{ode}\right) \tag{13}$$

$$\begin{aligned}\widetilde{E}_{fre} &= \alpha_{fre}\beta_{fre}E_{fre}\\ \widetilde{E}_{wie} &= \alpha_{wie}\beta_{wie}E_{wie}\\ \widetilde{E}_{ode} &= \alpha_{ode}\beta_{ode}E_{ode}\end{aligned} \tag{14}$$

$$AttE = concat\left(\widetilde{E}_{fre},\widetilde{E}_{wie},\widetilde{E}_{ode}\right) \tag{15}$$

In this context, the *MA* and *EA* are the modality attention and embedding attention, respectively. The $\alpha_{\bullet}$ is a scalar representing the modality weight for each input, while $\beta_{\bullet}$ is the embedding weight with the same shape $E_{\bullet}$. The refined embedding of each modality is obtained by the element-wise product of the input and weight (modality and embedding) in a parallel manner. The final output embedding is formulated by concatenating the attentive embedding for each modality, which will be fed into the PN network for the prediction task.

3.3.4. The ODE and PN module

In the proposed model, the ODE module is to generate an embedding from the flight-level operation details. Since the input is organized as a one-hot vector, the ODE module is fully built based on the FC layers. As shown in (16), X_o is input flight-level operation details described in Table 1, $SFC(\bullet)$ denotes the inference rules of the stacked FC layers, with the ReLU as the activation function for each layer.

$$E_{ode} = SFC(X_o) \tag{16}$$

Finally, the PN network serves as a predictor to map the abstract high-level representations into the expected fuel consumption for a certain flight. As in (17), the FC layer takes the embedding vector generated by the AttFN module as input, followed by a Sigmoid activation function to predict the output fuel consumption.

$$FCP = Sigmoid(\mathrm{FC}(AttE)) \tag{17}$$

3.4. Model output

As illustrated before, instead of a single scalar for denoting the fuel consumption of a certain flight, in this work, binary encoding is proposed to represent the final output using the binary numerical system. In this work, the integer-based fuel consumption is converted into the binary, which serves as the model output for the training and evaluation. The motivation of the BE representation is to prevent a sharp dimension reduction from the embedding feature to a scalar output (e.g., from 64 to 1), which further benefits to achieving an efficient and effective model convergence. By analyzing the fuel consumption in the dataset, a total of 16-bit binary is applied to denote the output, in which each bit can only be 0 or 1. For instance, the scalar 12,600 is denoted by "0011 0001 0011 1000″ using the BE representation.

Intuitively, by outputting the BE values, the traditional regression task is not suitable for this framework due to its data attributes (from float to binary). As shown in Fig. 3, for the integer-based representation, the output of the proposed model is a scalar, in which the training loss can be easily measured by regression-related factors for a single digit. For the BE representation, the outputs of the proposed model are sequential bits with different weights, in which each bit can only be 0 or 1. The prediction task of the proposed model is to determine the value of each bit, by classifying the outputs into the two available classes, i.e., the binary classification task. Finally, for the whole BE representation, the training loss is measured by combining all the binary classification losses in a bitwise manner.

To this end, an innovatively mBC mechanism is proposed to address this issue by redefining the FCP task as a multi-binary classification task. The goal of the model training is to predict the binary code (0 or 1) for each bit, which can be implemented by multiple binary classification tasks. To conduct the error measurement, the sigmoid activation is applied to the final layer to ensure that all the outputs are from 0 to 1. A bit less than 0.5 is regarded as binary 0, otherwise, set to binary 1.

3.5. Training loss

By defining the FCP task as multiple binary classification tasks, the binary cross-entropy is naturally selected as the loss function. The mBC loss can be illustrated as (18), in which K is the total bit of the BE representation, 16 in this work. BCE denotes the binary cross-entropy loss. The y_k^b and $\widehat{y}_k^b$ are the BE of the ground truth and the prediction for a certain sample, respectively.

$$\lambda_C = \sum_{k=1}^{K} BCE\left(y_k^b,\widehat{y}_k^b\right) \tag{18}$$

Considering the attributes of the binary numerical system, the weights of different bits make great contributions to the loss evaluation, as shown below:

For instance, the weight of the highest bit is 2^{15}, while the lowest bit is only 2^0. Based on the mBC loss, the λ_C between the "0011 0001 0011 1000″ (decimal in 12,600) and "1011 0001 0011 1000″ (decimal in 45,368) is the same as that between the "0011 0001 0011 1000″ (decimal in 12,600) and "0011 0001 0011 1001″ (decimal in 12,601), i.e., 1.7269.

However, their corresponding integer-based errors are $32{,}768^2$ and 1, respectively, which shows a large error deviation in this context. This fact will inevitably confuse the model training, and further impact the model performance.

Based on the aforementioned analysis, the Mean Squared Error (MSE) is also selected as a part of the loss function to prevent high-bit prediction errors. As listed in (19), the y and $\widehat{y}$ are the decimal of the ground truth and the prediction for a certain sample, respectively.

$$\lambda_R = MSE(y-\widehat{y}) \tag{19}$$

Finally, the loss of the proposed FCP model is defined as:

$$\lambda = \lambda_C + \omega\lambda_R \tag{20}$$

where the ω is a balance factor between the two kinds of loss, and we

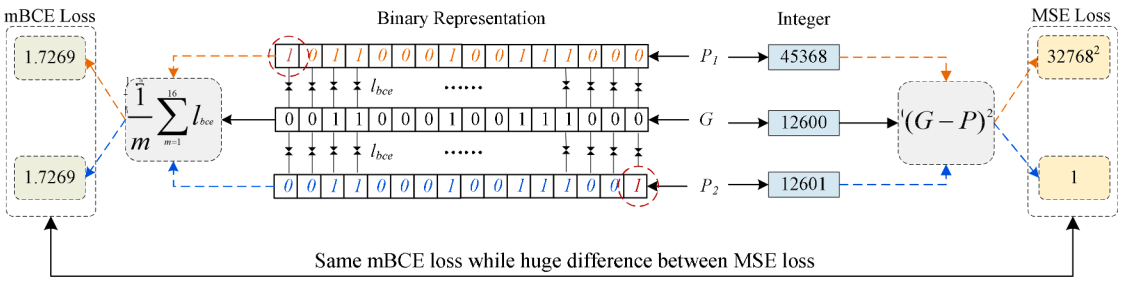

Fig. 3. The comparison between BCE and MSE loss.

manually set it to 0.001 based on generated loss during model training. The final loss considers both the classification loss and the typical regression loss to achieve the FCP task, which is expected to enhance the model convergence and finally improve the prediction performance.

4. Experimental configurations

4.1. Data descriptions

To validate the proposed approach, a real-world dataset is built by preparing multi-source information from air traffic control (ATC) centers, airlines, etc. The data period is from 2017/01/01 to 2017/12/31. The airline records relate to the Hong Kong-based flights, consisting of the actual departure and arrival time, fuel loadings, and actual fuel consumption of each flight, etc. A total of three OD pairs are selected as the cases in this work, involving Chengdu (ZUUU), Beijing (ZBAA), and Shanghai (ZSPD), which are representative airports in China.

The required data items of the flight route were collected from a real national ATC system, including the operation details. The resolution of the gridded wind velocity is 0.25° for the horizontal plane (longitude and latitude), while in the altitude dimension, the raw data is pre-processed to match the standard flight levels published by the Civil Aviation Administration of China (CAAC). The time interval of the weather information is 10-minute. The airport level is categorized by related CAAC regulations, i.e., a total of 45 flight levels in China.

As to the flight-related information, the nodes of the graph based on flight routes for the ZUUU, ZBAA, and ZSPD are 20, 22, and 16, respectively. The average flight time for the ZUUU, ZBAA, and ZSPD are about 166, 218, and 165 minutes, respectively. As to fleet mix for the flight operation, since the three OD pairs are median-distance airlines, the main models are Airbus aircraft, including A320, A321, and A33* series (such as 33C, 33H, and 33R). To simplify the aircraft classes, we categorize the fleet mix into A32* and A33* series, which are further applied to evaluate the prediction performance based on different aircraft models. The number of A32* models for ZUUU, ZBAA, and ZSPD are 48.1%, 22.5%, and 47.2%, respectively, while the remaining flights are operated on the A33* model. It can be seen that the A33* model is a preferred option for flights between ZBAA and VHHH due to the higher traffic demands in Beijing, China. In general, the average flight time for A33* aircraft is longer than that of the A32* model. Based on aforementioned descriptions, a sample data is presented as below:

DATE: 2017/01/01; FLIGHT: HDA825; ADEP: ZUUU; ADES: VHHH; AT: 321; TIME_DEP:00_11; TIME_ARR:02_40; ROUTE: ZUUU/CTU/BHS/ZYG/…/VIBOS/ZUH/VHHH; AFC: 5900.

Except the notations in Table 1, the TIME_DEP and TIME_ARR denote the actual time of the flight departure and arrival, respectively. The ROUTE saves the waypoint sequence split by "/", whose detailed information can be extracted from base data of the airport. The AFC is the actual fuel consumption (in kilogram) for the flight operation.

Some noisy samples are removed from the raw dataset due to missing items, outliers, etc. All items of the input data are normalized by the min-max operation based on their attributes. After the pre-processing steps, a total of 684, 2305, and 3457 samples are for ZUUU, ZBAA, and ZSPD, respectively. All the samples for each flight are randomly divided into two subsets to conduct the comparative experiments. Firstly, limited by the total number of samples in the dataset, we randomly select 90% of the total samples to construct the training dataset to ensure sufficient training samples in this context. The rest 10% of the samples are regarded as the evaluation dataset to report the model performance. No validation dataset is designed due to limited samples in this dataset. Alternatively, the training procedure is performed with fixed epochs to optimize the proposed model. The trainable hyper-parameters are saved at the end of each epoch for further evaluation. In this work, all the reported metrics are tested on the evaluation dataset (10%) using the saved model parameters.

4.2. Implementation details

The proposed model is implemented using the open-source deep learning framework PyTorch 1.3.1. The training server is equipped with 2 Intel Core i7–6800 K processors, 2 NVIDIA GeForce GTX 2080Ti GPUs, 128-GB memory, and an Ubuntu 16.04 operating system.

During the model training, the Adam optimizer is applied to fit the model parameters. The initial learning rate is 0.01 and will be multiplied by 0.9 every 20 epochs. The batch size is set to 16. All the training samples are shuffled to improve model robustness. The total training epochs are set to 100 to perform the training procedure. Due to limited training samples, we attempt to design a lightweight and efficient architecture to enhance the prediction performance and prevent the overfitting problem. In the early stage, we investigate the influence of different hyper-parameter options on determining prediction performance. The resulting configurations of the proposed architecture are summarized as follows:

A total of two GCN layers with 16 neurons are configured to construct the FRE module. For the WIE module, a total of 16 filters are for 3D-Conv, and the kernel size is (3, 3, 2) with a stride of 3. The input length of the ConvLSTM operations covers flight duration and is determined by the update interval of the weather information. The convolution operation in the ConvLSTM block is with 16 filters, kernel size of (3, 3), and stride of 3. The final ConvLSTM block only generates the features of the last step to support the following embedding. For the ODE module, a total of two FC layers are designed, with 32 and 16 neurons. The dimension of the embedding vector for the FRE, WIE and ODE modules are 64, 64, and 16, respectively. For the AttFN module, two hidden FC layers with 4 neurons are designed for modality attention, while hidden FC layers with 64 neurons are for embedding attention. The output dimension of the PN network is the bit number of the BE representation, i.e., 16 in this work.

4.3. Evaluation metrics

The performance evaluation is conducted based on the decimal-based fuel consumption between the prediction and ground truth. The following metrics, including the root mean squared error (RMSE), mean

absolute error (MAE), and MAPE, are selected to evaluate the model performance:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \widehat{y}_i)^2} \tag{21}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|(y_i - \widehat{y}_i)| \tag{22}$$

$$MAPE = 100\% \times \left(\frac{1}{m}\sum_{i=1}^{m}\left|\frac{(y_i - \widehat{y}_i)}{y_i}\right|\right) \tag{23}$$

where m is the number of test samples, and the y_i and $\widehat{y}_i$ are the real and prediction result for the $i-th$ sample, respectively. Since the fuel loading decision is safety-critical for airline operations, the risk indicator (RI) is also introduced to evaluate the applicability of FCP approaches. The equation is shown below, where rf_i denotes the reserve fuel for the $i-th$ sample.

$$RI = \frac{1}{m}\sum_{i=1}^{m}R_i, R_i = \begin{cases} 1 & (\widehat{y}_i - y_i)\langle rf_i \\ 0 & else \end{cases} \tag{24}$$

4.4. Baseline models

In this work, several baselines are also designed to further confirm the performance superiority of the proposed model. Based on the task specificities, the regression approaches are selected as baseline models. The primary criteria are to confirm the technical improvements of the proposed approach, considering the time series (RNN-based) and graph (GCN-based) nature of the flight route and other required information using deep architecture. Learning from related works in this work, the following baselines are selected to conduct the comparative experiments, as shown below:

a) LASSO [35]: least absolute shrinkage and selection operator. As in [4], it is a commonly used method for regression with high dimensional features, which is formulated based on least squares models with L1 regularization.
b) DNN [6]: i.e., deep neural networks-based model. FC layers are applied to construct this baseline. All the inputs are encoded into a 1D vector, including the flight route, weather information, and operational details. A total of 5 hidden layers are designed with 128, 64, 32, 16, and 8 neurons.
c) GNN [26]: GNN layers are applied to build this model, whose original input is the same as the proposed approach. The weather information is also embedded into the input feature by averaging the 3D grid data based on the waypoint positions.
d) LSTM [27]: the core block is LSTM in this baseline, in which the input features are the same as that of the GNN model. A total of two LSTM layers are designed with 16 and 8 neurons, where an FC layer serves as a predictor.

5. Results and discussions

5.1. Overall performance

With the aforementioned experimental configurations, the proposed approach and baselines are trained and evaluated on the same test dataset. The performance in terms of given measurements is reported in Table 2.

From the experimental results, we can see that the proposed approach achieves the best performance among all the approaches. In general, thanks to the ability to model the non-linear features, the neural architectures yield higher prediction accuracy. The LASSO model suffers from the poorest performance due to the complex patterns of the high-dimensional input features for the FCP task. As to the deep learning models, the DNN model fails to obtain the desired performance since no certain blocks are explicitly designed to capture the spatial and temporal correlations. The results obtained by the two approaches are also inferior to the real fuel prediction from the current system.

Table 2
Performance of different approaches.

	RMSE	MAE	MAPE (%)	RI
Real[1]	1322.4	1021.5	7.76	0.000
LASSO	1526.5	1228.7	10.26	0.035
DNN	1450.7	1159.4	9.75	0.029
LSTM	1335.3	1049.7	8.13	0.021
GNN	1290.2	1012.9	7.61	0.018
FCPNet*	**1053.4**	**837.1**	**6.50**	**0.009**

[1] The results are obtained by collecting data items from tools of current airline operation.

As can be seen from the results, the LSTM and GNN models reach higher accuracy due to their designs for learning temporal and spatial patterns, respectively. Specifically, the GNN model obtains better performance, which proves that the graph spatial correlations play a more important role in predicting fuel consumption based on the flight route (rather than temporal dependencies). Finally, the proposed model harvests the best performance, with only 6.50% MAPE for the pre-departure fuel loading decision. It can be attributed that dedicated modules are designed for the proposed model to capture required underlying patterns from the input features to the fuel consumption.

In this experiment, we also consider the prediction accuracy for different OD flight pairs to check the model generalization. As shown in Fig. 4, the distributions of predicted MAPE measurements for each OD pair are organized as a boxplot, in which they are generally similar to each other, indicating that the proposed model can be properly generalized to different cases for further improving its applicability. It can also be seen that compared to the flights from ZUUU, the flight pairs departing from the ZBAA and ZSPD suffers from larger prediction error, which can be attributed that the fluctuance factors of the flight flow caused by busy traffic situations impact the transition patterns of the FCP task in the two airports and corresponding flight routes. As analyzed from the flight schedule, the ZSPD airport is a trunk base for most international flights, which injects more uncertainties (outside the China mainland) into the traffic system and hence reduces the model accuracy. To be specific, we further investigate the underlying causes of the prediction results, as shown as follows:

a) It can be seen that some outliners are in the prediction results, as shown in the stars in Fig. 4. By considering the input information, we find that almost all the outliner predictions are caused by flight re-routing with different flight environments, such as the flight profile, weather information, etc. In the FCP task, the flight re-routing is expected to result in minor samples in the dataset, which hence degrades the performance of the data-driven models. In addition, they are also the primary contributors to inducing RI.
b) Furthermore, the results located in the upper quartile in the boxplot are also considered to analyze the main contributors to the performance deviation. Based on the detailed information of the flight operation, we find that the rerouting, weather information, flight time and are the top-3 factors for the larger prediction error. This result also confirms the motivation of feeding the multi-modality inputs into the proposed model, i.e., flight route, weather information, and operational details.

In addition, the MAPE measurements for different OD pairs categorized by aircraft models are also considered to evaluate the prediction performance. As shown in Fig. 5, for all three OD pairs, the prediction performance for the A32* aircraft model is higher than that of the A33*. In general, flight operation with A33* indicates higher traffic demands

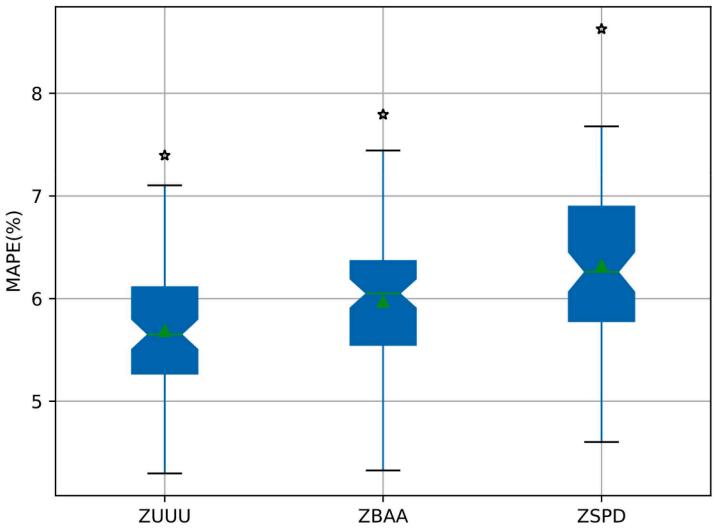

Fig. 4. The distribution flight-wise prediction accuracy for different OD pairs.

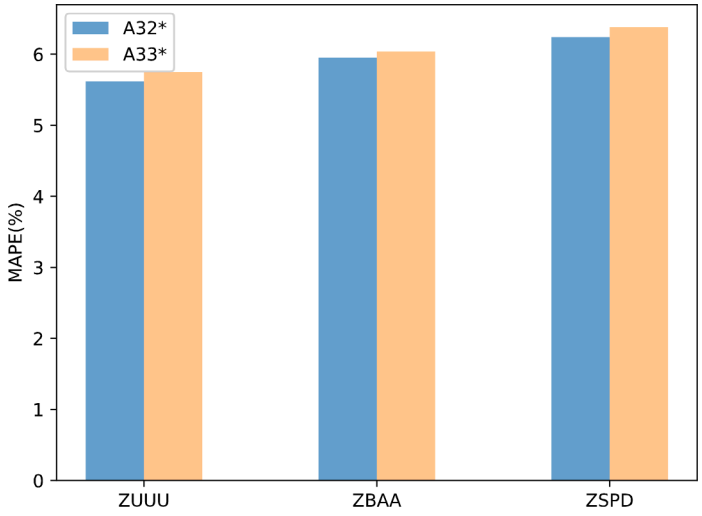

Fig. 5. The performance for different OD pairs categorized by aircraft models.

for certain OD pairs, which incorporates more uncertainties for the prediction task, such as aircraft performance, flight profile, etc. As analyzed from the raw data, we found that the gaps of the fuel consumption between the A32* and A33* for all OD pairs are about 2%, which also validates the motivation for considering the multi-modality inputs in this work.

5.2. Ablation study

5.2.1. Validation for architecture improvements

In this section, ablation studies are also conducted to focus on the efficacy of different modules for processing multi-modal information, which is expected to provide a better understanding. The experiments are conducted based on the paths from each basic module to the proposed model, in which two primary categories (with and without AttFN) are designed to validate each module, as shown in Table 3.

As to experiments 1–4, different combinations of multi-modality

Table 3
Performance of ablation studies.

No.	FRE	WIE	ODE	AttFN	RMSE	MAE	MAPE (%)	RI
1	✓				1300.9	1016.2	7.81	0.021
2	✓	✓			1233.4	990.7	7.50	0.015
3	✓		✓		1284.5	1004.3	7.72	0.016
4	✓	✓	✓		1171.4	940.0	7.21	0.010
5	✓			✓	1279.9	1003.5	7.73	0.018
6	✓	✓		✓	1181.1	946.9	7.26	0.010
7	✓		✓	✓	1254.7	978.2	7.50	0.015
8	✓	✓	✓	✓	**1053.4**	**831.1**	**6.50**	**0.009**

information are fed into the proposed model. The FRE module serves as a baseline, which is similar to the GNN model with only flight route information. From the results, we can see that each module provides desired performance improvements, and the prediction MAPE reduces from 7.81% to 7.21%. To be specific, based on the essential input (i.e., FRE), the WIE obtains higher accuracy than that of the ODE module, benefiting from the significant influence of weather information on aircraft performance during flight operation. Most importantly, the proposed model with only the FRE module has the ability to outperform some baselines, which indicates that the proposed model can achieve the FCP task with the desired performance solely based on planned route information.

The best performance can be obtained by combining all the three modules proposed in this work (with multi-modality information), which also achieves a higher RI performance. With the same input information, experiments 2 and 3 achieve higher prediction accuracy than that of the GNN baseline (as shown in Table 2), which indicates that the explicit designs for capturing spatial and temporal correlations provide significant contributions to enhancing the final FCP performance.

From experiments 5–8, the AttFN module is further validated by combining it with different input combinations. It can be seen that the AttFN can contribute accuracy improvements for all the input combinations. Specifically, more improvements can be obtained with the increasing of the input modality, and thus the AttFN only yields marginal improvement with the FRE module. The results in experiments 6 and 7 share similar trends with those of experiments 2 and 3, i.e., the WIE module harvests higher performance improvement than that of the ODE module.

Finally, the most prominent accuracy improvement resulting from the AttFN module is reached by the proposed model with all three kinds of input information, i.e., about 0.71% absolute MAPE gains. The results confirm the effectiveness of the attention mechanism for selecting required task-oriented features from a cluster of intermediate feature maps.

Based on the aforementioned ablation studies, it can be concluded that all the proposed modules play a desired role in improving the final accuracy, which supports the motivation of the proposed model, i.e., considering multi-modality inputs by leveraging the attention mechanism to learn informative features with dedicated weights.

5.2.2. Validation for BE and loss function

In this section, we mainly focus on validating the proposed BE representation and related loss combination. In the designed experiments, the full model architecture is applied to achieve the FCP task, with only the modification of the model output. The results are reported in Table 4, in which the two categories are listed, i.e., the integer output is the typical regression problem, while the BE output is a classification problem.

As can be seen from the results that the BE representation achieves higher performance than that of the integer representation, in terms of both four measurements. As to the BE representation, a total of three experiments are conducted to confirm the final performance with different loss functions. As reported in the results, the proposed loss combination (BCE and MSE) achieves higher performance compared to the results obtained by a single loss. The BE results based on MSE loss are not reported here since we fail to achieve a model convergence, which can be attributed to the difference between the mBC mechanism (classification task) and the MSE loss (regression task).

It is also noted that the BE representation with only BCE loss suffers an inferior RI, which is important to the safety-critical task in this work. By analyzing the experimental results, we found that the incorrect predictions on the high-weighted bits are the primary contributor to the larger RI, which cannot be optimized by the classification-only loss function, as illustrated before. In conclusion, the results with larger RI also support our motivation to apply a combined loss of (BCE and MSE), which is also an indispensable design for the FCP task.

Table 4
Performance of BE and loss function.

Output	Loss	RMSE	MAE	MAPE (%)	RI
Integer	MSE	1071.8	840.4	6.57	0.009
BE	BCE	1064.6	837.4	6.54	0.012
	MSE	–	–	–	–
	Proposed*	**1053.4**	**831.1**	**6.50**	**0.009**

5.3. Discussions

5.3.1. The weather impact

Once a flight takes off, the weather near the planned route plays a decisive role in the future flight operation. In this section, we further analyze the weather impacts for the FCP task by visualizing the heatmap of the last layer in the WIE module. As shown in Fig. 6, the activations of the feature map in different areas are visualized using third-party tools. Except for some noisy points, the proposed model mainly focuses on the route of the flight OD pair, indicating that the weather changes in those areas contribute higher impacts on the fuel consumption for flight operation. This also confirms the efficacy of the attention mechanism for learning informative patterns from certain input features or patches. In general, the learned attention weights are particularly higher around the departure and landing airport, which can be attributed to that for the actual flight operation, most congestions and intervene are occurred in terminal airspaces (near the airport) due to dense traffic situations. The mentioned traffic operations may lead to a flight delay or re-route, and finally impact the accuracy of the FCP task. It can also be concluded that a global traffic flow situation for certain airspaces should be considered to further improve the prediction accuracy.

5.3.2. Flight time

In this section, we consider the relationship between the total flight time and the prediction error in MAPE. As shown in Fig. 7, for a certain flight, a longer flight time possibly indicates higher uncertainties caused by real-time factors, such as traffic congestion or traffic flow control, which breaks the plannability and predictability, thus suffering from high errors. This trend is confirmed by all three cases, i.e., longer flight time suffers from larger prediction errors due to real-time unpredictable factors.

Meanwhile, among different flights, the proposed model harvests better performance for the OD pair with longer flight time, i.e., as the red line in Fig. 7, for the 170-minute flight time, the ZBAA flights obtain about 4% MAPE, while ZUUU and ZSPD flights yield only 7% MAPE. As well known, for a flight with a longer duration, its cruise time accounts for a higher percentage of the total time since the departure and landing time are generally consistent for all flights. Thus, the proposed model achieves higher performance since the fuel consumption for the stable motion state in the cruise phase is easy to predict.

5.3.3. Pre- and post-departure analysis

In this section, an extra experiment is conducted to compare the performance of the proposed approach (pre-departure) with a post-analysis approach. Unlike the pre-departure prediction in this work, the post-flight analysis achieves the FCP task after the flight operation based on real collected trajectories. The comparative approach is implemented by referring to [4].

The final experimental results are listed in Fig. 8. Compared to the baseline, the proposed model fails to provide a comparable accuracy for the prediction before flight departure, but it also can provide promising improvement based on the current airline practice. This fact mainly indicates that the comparative model can capture the required underlying patterns of the FCP task since it is driven by real collected trajectories, which currently serve as a post-event analysis tool.

Fortunately, the proposed approach has the ability to predict the fuel

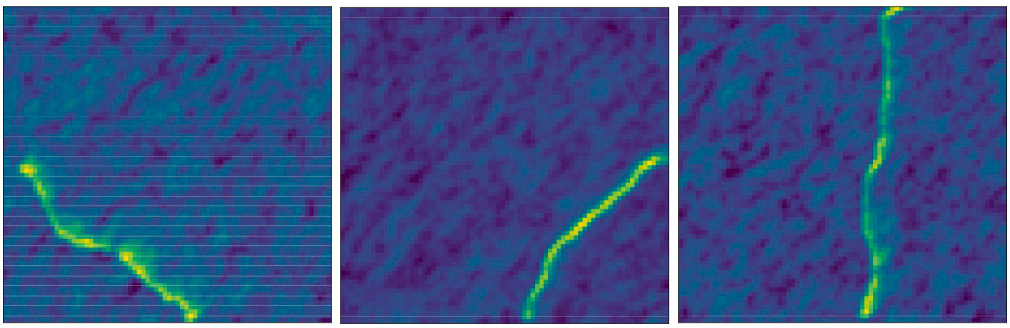

Fig. 6. The heatmap of the WIE module based on attention weights. (from left to right are ZUUU, ZSPD and ZBAA, respectively).

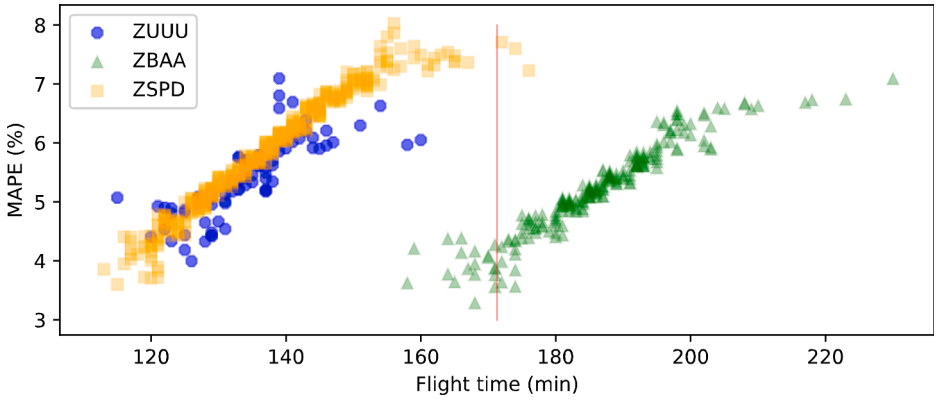

Fig. 7. Performance (MAPE) with different flight time.

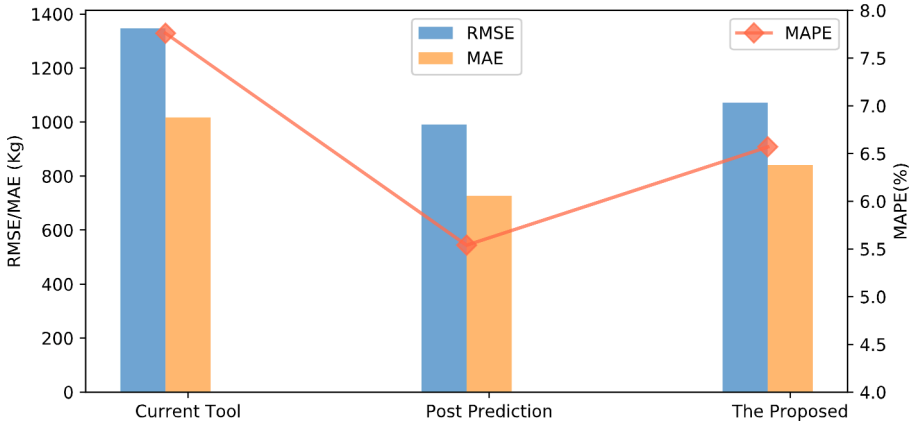

Fig. 8. Performance comparison for different approaches.

consumption before the flight departure based on the planned route and other required information, which practically supports the decision of the fuel loading for the airline operation. Based on the proposed approach, more proper fuel loading decisions can be made to control the "dead weight" of the flight, which further reduces the fuel consumption (also saves the airline cost) and achieves the global carbon-neutral target.

6. Conclusions

In this work, to reduce waste emissions and airline costs, a novel deep learning model, called FCPNet, is proposed to predict fuel consumption, which further supports the fuel loading decision before the flight departure. A real-world dataset is built to validate the proposed model by collecting multi-modality information from a national ATC system, airlines, etc. The results of extensive experiments demonstrate that the proposed model outperforms other baselines in terms of four measurements. In addition, the efficacy and effectiveness of the proposed multi-modality modules are also separately confirmed by several ablation studies. Most importantly, the proposed model is able to achieve acceptable performance for the pre-departure prediction task with only the planned route.

In the future, more advanced neural architectures also deserve to be considered to process multi-modality information. We also plan to introduce more operational information (such as flight delay, and the traffic flow) to enhance the model performance.

Funding

This work was supported by the National Natural Science Foundation

of China under Grant 62371323, 62001315 and U20A20161, the Hong Kong Research Grants Council General Research Fund under Grant 11215119, the Open Fund of Key Laboratory of Flight Techniques and Flight Safety, Civil Aviation Administration of China (CAAC) under Grant FZ2021KF04, and the Fundamental Research Funds for the Central Universities under Grant No. 2021SCU12050.

CRediT authorship contribution statement

Yi Lin: Conceptualization, Methodology, Software, Validation, Writing – original draft. **Dongyue Guo:** Data curation, Conceptualization, Validation. **Yuankai Wu:** Investigation, Validation, Writing – review & editing. **Lishuai Li:** Data curation, Writing – review & editing, Funding acquisition. **Edmond Q. Wu:** Investigation, Software, Validation, Supervision. **Wenyi Ge:** Data curation, Methodology, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] Khalili Siavash, Rantanen Eetu, Bogdanov Dmitrii, Breyer Christian, Global Transportation demand development with impacts on the energy demand and greenhouse gas emissions in a climate-constrained world, Energies 12 (20) (2019) 3870, https://doi.org/10.3390/en12203870.
[2] Capoccitti Sam, Khare Anshuman, Mildenberger Udo, Aviation industry - mitigating climate change impacts through technology and policy, J. Technol. Manag. Innov. 5 (2) (2010), https://doi.org/10.4067/S0718-27242010000200006.
[3] Gössling Stefan, Humpe Andreas, Fichert Frank, Creutzig Felix, COVID-19 and pathways to low-carbon air transport until 2050, Environ. Res. Lett. 16 (3) (2021), https://doi.org/10.1088/1748-9326/abe90b.
[4] Zhu Xinting, Li Lishuai, Flight time prediction for fuel loading decisions with a deep learning approach, Transp. Res. Part C Emerg. Technol. 128 (2021), https://doi.org/10.1016/j.trc.2021.103179.
[5] Kang Lei, Hansen Mark, S. Ryerson Megan, Evaluating predictability based on gate-in fuel prediction and cost-to-carry estimation, J. Air Transp. Manag. 67 (2018), https://doi.org/10.1016/j.jairtraman.2017.11.006.
[6] Uzun Mevlut, Demirezen Mustafa Umut, Inalhan Gokhan, Physics guided deep learning for data-driven aircraft fuel consumption modeling, Aerospace 8 (2) (2021), https://doi.org/10.3390/aerospace8020044.
[7] S. Ryerson Megan, Hansen Mark, Bonn James, Time to burn: flight delay, terminal efficiency, and fuel consumption in the national airspace system, Transp. Res. Part A Policy Pract. 69 (2014), https://doi.org/10.1016/j.tra.2014.08.024.
[8] K. Seymour, M. Held, G. Georges, K. Boulouchos, Fuel estimation in air transportation: modeling global fuel consumption for commercial aviation, Transp. Res. Part D Transp. Environ. 88 (2020), https://doi.org/10.1016/j.trd.2020.102528.
[9] Currie Charlotte, Marcos Andres, Turnbull Oliver, Wind optimal flight trajectories to minimise fuel consumption within a 3 dimensional flight network, in: 2016 UKACC 11th Int. Conf. Control, 2016.
[10] Lin Yi, Zhang Jian-wei, Liu Hong, Deep learning based short-term air traffic flow prediction considering temporal–spatial correlation, Aerosp. Sci. Technol. 93 (2019), 105113, https://doi.org/10.1016/j.ast.2019.04.021.
[11] Singh Vedant, Sharma Somesh Kumar, Fuel consumption optimization in air transport: a review, classification, critique, simple meta-analysis, and future research implications, Eur. Transp. Res. Rev. 7 (2) (2015) 12, https://doi.org/10.1007/s12544-015-0160-x.
[12] S. Ryerson Megan, Hansen Mark, Hao Lu, Seelhorst Michael, Landing on empty: estimating the benefits from reducing fuel uplift in US Civil Aviation, Environ. Res. Lett. 10 (9) (2015), https://doi.org/10.1088/1748-9326/10/9/094002.
[13] Horiguchi Yuji, Baba Yukino, Kashima Hisashi, Suzuki Masahito, Kayahara Hiroki, Maeno Jun, Predicting fuel consumption and flight delays for low-cost airlines, in: Proc. Twenty-Ninth AAAI Conf. Innov. Appl., San Francisco, California, USA, 2017, pp. 4686–4693.
[14] Trani Antonio, F. Wing-Ho, Schilling Glen, Baik Hojong, Seshadri Anand, A neural network model to estimate aircraft fuel consumption, in: AIAA 4th Aviat. Technol. Integr. Oper. Forum, Reston, Virigina, 2004.
[15] S. Chati Yashovardhan, Balakrishnan Hamsa, Data-driven modeling of aircraft engine fuel burn in climb out and approach, Transp. Res. Rec. J. Transp. Res. Board 2672 (29) (2018), https://doi.org/10.1177/0361198118780876.
[16] Woodbury Timothy, Srivastava Ashok, Analysis of Virtual Sensors For Predicting Aircraft Fuel Consumption, Infotech@aerosp, Reston, Virigina, 2012, p. 2012.
[17] Huang Chenyu, Cheng Xiaoyue, Estimation of aircraft fuel consumption by modeling flight data from avionics systems, J. Air Transp. Manag. 99 (2022), 102181, https://doi.org/10.1016/j.jairtraman.2022.102181.
[18] Kang Lei, Hansen Mark, Improving airline fuel efficiency via fuel burn prediction and uncertainty estimation, Transp. Res. Part C Emerg. Technol. 97 (2018), https://doi.org/10.1016/j.trc.2018.10.002.
[19] Yao Zhihong, Wang Yi, Liu Bo, Zhao Bin, Jiang Yangsheng, Fuel consumption and transportation emissions evaluation of mixed traffic flow with connected automated vehicles and human-driven vehicles on expressway, Energy 230 (2021), 120766, https://doi.org/10.1016/j.energy.2021.120766.
[20] Yao Zhihong, Deng Haowei, Wu Yunxia, Zhao Bin, Li Gen, Jiang Yangsheng, Optimal lane-changing trajectory planning for autonomous vehicles considering energy consumption, Expert Syst. Appl. 225 (2023), 120133, https://doi.org/10.1016/j.eswa.2023.120133.
[21] Lin Yi, Li Linchao, Jing Hailong, Ran Bin, Sun Dongye, Automated traffic incident detection with a smaller dataset based on generative adversarial networks, Accid. Anal. Prev. 144 (2020), 105628, https://doi.org/10.1016/j.aap.2020.105628.
[22] Lin Yi, Wu YuanKai, Guo Dongyue, Zhang Pan, Yin Changyu, Yang Bo, et al., A deep learning framework of autonomous pilot agent for air traffic controller training, IEEE Trans. Human-Machine Syst. 51 (5) (2021) 442–450, https://doi.org/10.1109/THMS.2021.3102827.
[23] Wu Yuankai, Tan Huachun, Qin Lingqiao, Ran Bin, Jiang Zhuxi, A hybrid deep learning based traffic flow prediction method and its understanding, Transp. Res. Part C Emerg. Technol. 90 (2018) 166–180, https://doi.org/10.1016/j.trc.2018.03.001.
[24] Liu Hong, Lin Yi, Chen Zhengmao, Guo Dongyue, Zhang Jianwei, Jing Hailong, Research on the air traffic flow prediction using a deep learning approach, IEEE Access 7 (2019) 148019–148030, https://doi.org/10.1109/ACCESS.2019.2945821.
[25] Jepsen Tobias Skovgaard, S. Jensen Christian, Nielsen Thomas Dyhre, Relational fusion networks: graph convolutional networks for road networks, IEEE Trans. Intell. Transp. Syst. (2020), https://doi.org/10.1109/TITS.2020.3011799.
[26] Zhao Ling, Song Yujiao, Zhang Chao, Liu Yu, Wang Pu, Lin Tao, et al., T-GCN: a temporal graph convolutional network for traffic prediction, IEEE Trans. Intell. Transp. Syst. 21 (9) (2020), https://doi.org/10.1109/TITS.2019.2935152.
[27] Shi Zhiyuan, Xu Min, Pan Quan, Yan Bing, Zhang Haimin, LSTM-based flight trajectory prediction, in: 2018 Int. Jt. Conf. Neural Networks, 2018, pp. 1–8.
[28] Jaafar Noussaiba, Lachiri Zied, Multimodal fusion methods with deep neural networks and meta-information for aggression detection in surveillance, Expert Syst. Appl. 211 (2023), 118523, https://doi.org/10.1016/j.eswa.2022.118523.
[29] Yu Wentao, Zeiler Steffen, Kolossa Dorothea, Fusing information streams in end-to-end audio-visual speech recognition, in: ICASSP 2021 - 2021 IEEE Int. Conf. Acoust. Speech Signal Process, 2021, pp. 3430–3434.
[30] Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, N. Gomez Aidan, et al., Attention is all you need, Adv. Neural Inf. Process. Syst. (2017).
[31] Hu Jie, Shen Li, Albanie Samuel, Sun Gang, Wu Enhua, Squeeze-and-Excitation Networks, 2017.
[32] Woo Sanghyun, Park Jongchan, Lee Joon-Young, Kweon in so. CBAM: convolutional block attention module, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) (2018) 3–19.
[33] Shi Xiaoming, Qi Heng, Shen Yanming, Wu Genze, Yin Baocai, A spatial–temporal attention approach for traffic prediction, IEEE Trans. Intell Transp. Syst. 22 (8) (2021), https://doi.org/10.1109/TITS.2020.2983651.
[34] Xue Zihui, Marculescu Radu, Dynamic multimodal fusion, in: Multi-Modal Learn. Appl. Work. (MULA), CVPR, 2023, pp. 2574–2583.
[35] Pedregosa Fabian, Varoquaux Gaël, Gramfort Alexandre, Michel Vincent, Thirion Bertrand, Grisel Olivier, et al., Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.