Aircraft Bleed Air System Fault Prediction based on Encoder-Decoder with Attention Mechanism

Indexed by:
Web of Science Group

Siyu Su[a], Youchao Sun[a,*], Chong Peng[a], Yifan Wang[a]

[a] Nanjing University of Aeronautics and Astronautics, China

Highlights

- A novel fault prediction method for the aircraft bleed air system is proposed by combining the DSTP-ED prediction model and the EWMA control chart.

- The DSTP-ED model incorporates attention mechanisms and has more accurate prediction results compared to other models.

- The EWMA control chart can effectively identify impending bleed air system failures.

- The proposed method is validated with real airline QAR data.

Abstract

The engine bleed air system (BAS) is one of the important systems for civil aircraft, and fault prediction of BAS is necessary to improve aircraft safety and the operator's profit. A dual-stage two-phase attention-based encoder-decoder (DSTP-ED) prediction model is proposed for BAS normal state estimation. Unlike traditional ED networks, the DSTP-ED combines spatial and temporal attention to better capture the spatiotemporal relationships to achieve higher prediction accuracy. Five data-driven algorithms, autoregressive integrated moving average (ARIMA), support vector regression (SVR), long short-term memory (LSTM), ED, and DSTP-ED, are applied to build prediction models for BAS. The comparison experiments show that the DSTP-ED model outperforms the other four data-driven models. An exponentially weighted moving average (EWMA) control chart is used as the evaluation criterion for the BAS failure warning. An empirical study based on Quick Access Recorder (QAR) data from Airbus A320 series aircraft demonstrates that the proposed method can effectively predict failures.

Keywords

bleed air system, fault prediction, attention mechanism, deep learning, EWMA control chart

1. Introduction

The engine bleed air system (BAS) is a type of aircraft air source system that provides compressed air with regulated pressure and temperature for user systems (engine starting, air conditioning, wing ice protection, hydraulic reservoir pressurization, and pressurized water). BAS failure has a greater impact on the flight of the aircraft, which may lead to the abnormality of the cabin pressurization system, the performance degradation of the air conditioning system, etc. If the double-engine BAS failure occurs, it will often lead to homeward flight. Advanced failure prediction technology is urgently needed to reduce aircraft operation and maintenance costs due to performance degradation and system failures. Predictive maintenance with early identification of BAS malfunctions can reduce the number of aircraft stops and reduce airline operating costs.

BAS failure prediction can be achieved by a modeling approach based on systematic principles and a data-driven approach. But building a sufficiently accurate analytical model based on system principles is challenging, especially for nonlinear and complex systems. With the increasing availability of system monitoring data, data-based techniques have become an essential complement to model-based methods for fault prediction and diagnosis [10, 22, 40]. The data-driven approach

(*) Corresponding author.
E-mail addresses: S. Su (ORCID: 0000-0002-5390-2898) siyu_su@nuaa.edu.cn, Y. Sun (ORCID: 0000-0002-8358-8019) sunyc@nuaa.edu.cn, C. Peng chong_peng@nuaa.edu.cn, Y. Wang wang_yf@nuaa.edu.cn

simply uses the collected operational data to derive the model without requiring much knowledge of the system degradation mechanisms. Currently, data-driven fault prediction methods have been widely used in industrial systems and are classified into statistical methods (e.g., autoregressive models [36], statistical process techniques [21, 33], and mathematical morphology spectrum entropy [39], etc.) and machine learning (such as neural networks [26, 41], support vector machines [34, 37], and fuzzy methods [8, 35]). Aircraft have accumulated a large amount of sensor monitoring data during the operation, which can be classified as FDR (Flight Data Recorder), QAR (Quick Access Recorder), and ACARS (Aircraft Communications Addressing and Reporting System) data [30], according to the type of records. These data can reflect the operational status of aircraft systems and can be used to build models for condition monitoring, fault detection, etc. In the field of commercial aviation, predictive analytics for the exponentially growing number of operations and maintenance data generated on aircraft is full of promise [13, 38, 42]. However, there are limited studies on BAS in the published literature. Shang et al. [27, 28] developed a fault detection method for BAS's temperature sensors and valve actuators. Abdelrahman et al. [1] used backpropagation algorithms to model the faults of the most important bleed air system valves of the B-737 aircraft under desert conditions. Peltier et al. [23, 24] conducted an experimental investigation of the performance of different BASs. These studies focused on design improvements and component-specific troubleshooting while ignoring BAS operational data. Su et al. [31] established a risk warning model for BAS based on QAR data, but the adopted method did not consider the time-series relationship between the data, and there was a risk of masking failures.

The selection of feature variables, regression model, and pre-warning thresholds of the BAS fault prediction model all affect the results of fault prediction. And the main difficulty of BAS fault prediction research is selecting an effective prediction model to solve the BAS high-dimensional, nonlinear, and other fault prediction problems. The autoregressive integrated moving average model (ARIMA) [2] is a well-known method for temporal predictive analysis, with the advantage of a simple model that requires only endogenous variables and no other exogenous variables. However, ARIMA is only applicable to smooth time-series data, which can only capture linear relationships by nature, and the data used must be autocorrelated. Support vector regression (SVR) [18] is an application of support vector machines for regression problems. SVR is more concerned with the spatial correlation of data, which is effective in solving problems with high-dimensional features, but ignores the temporal correlation between data. The recurrent neural network (RNN) [9] is a class of neural networks for processing time-series data, where the current output of the series is related to the previous output. It is more suitable for short-term memory-type tasks and cannot address long-term dependency problems, and the gradient disappearance becomes more severe as the network complexity increases. Long short-term memory (LSTM) solves the gradient disappearance and explosion problem during long sequence training and is widely used in machine translation, speech recognition, and image processing [3, 11, 15]. The encoder-decoder (ED) framework was first introduced by Cho et al. [6] in sequence-to-sequence recurrent neural networks and is popular in machine translation [32]. However, since the context vector C of the encoder-decoder network is fixed, the model still does not work well for longer sequences [5]. To address this problem, Bahdanau et al. [3] proposed the attention mechanism, which has now been applied to natural language processing, time series prediction, etc [4, 7, 12]. Qing et al. [25] and Liang et al. [17] used a dual-stage attention-based RNN for time series forecasting, which solved the problem of long-term dependence but ignored the spatial association between the driving and target sequences.

Therefore, this paper adopts a dual-stage two-phase attention-based encoder-decoder model (DSTP-ED) for BAS fault prediction [14, 19], which can better capture the spatial-temporal relationships between data and performs well on various datasets. The "dual-stage" refers to using spatial and temporal attention mechanisms to obtain the spatiotemporal correlation between the driving and target sequences. The "two-phase" means that the spatial attention mechanism is composed of two attention modules for capturing the spatial correlation of the driving sequence and the target sequence. The main contributions of this paper are as follows:

(1) A data-driven failure prediction method is developed to monitor the condition of the BAS and identify its impending failures. The fault prediction method consists of a state

prediction model based on DSTP-ED and an anomaly criterion based on EWMA (Exponentially Weighted Moving Average) control chart. This organic combination of deep learning and statistical process control can effectively identify anomalies in BAS and detect faults in advance.

(2) The DSTP-ED model integrates spatial attention and temporal attention mechanisms that can adaptively select the most relevant input features as well as better capture long-term dependencies, effectively improving the accuracy of BAS state prediction. Compared with the four algorithms, ARIMA, SVR, LSTM, and ED, DSTP-ED has the optimal model performance.

(3) Fault prediction results may be disturbed due to changes in the operating environment and sensor noise. The EWMA control chart incorporates historical information from previous observations, which can well eliminate these anomalous effects

and reduce false alarms. Real aircraft operational data verify that EWMA detects BAS anomalies earlier and with fewer false alarms than other statistical methods.

The remaining sections of this paper are organized as follows: Section 2 presents the fault prediction theory in detail, including the DSTP-ED model, the training procedure, the fault diagnosis criteria, and the fault prediction process. Section 3 provides an application case of BAS fault prediction based on QAR data and presents the proposed method's comparative analysis. The conclusion is shown in the final section.

2. Fault prediction theory

This section describes the framework for fault prediction and the detailed algorithms, specifically the DSTP-ED prediction model and the fault monitoring methods. The fault prediction framework of BAS is shown in Fig. 1.
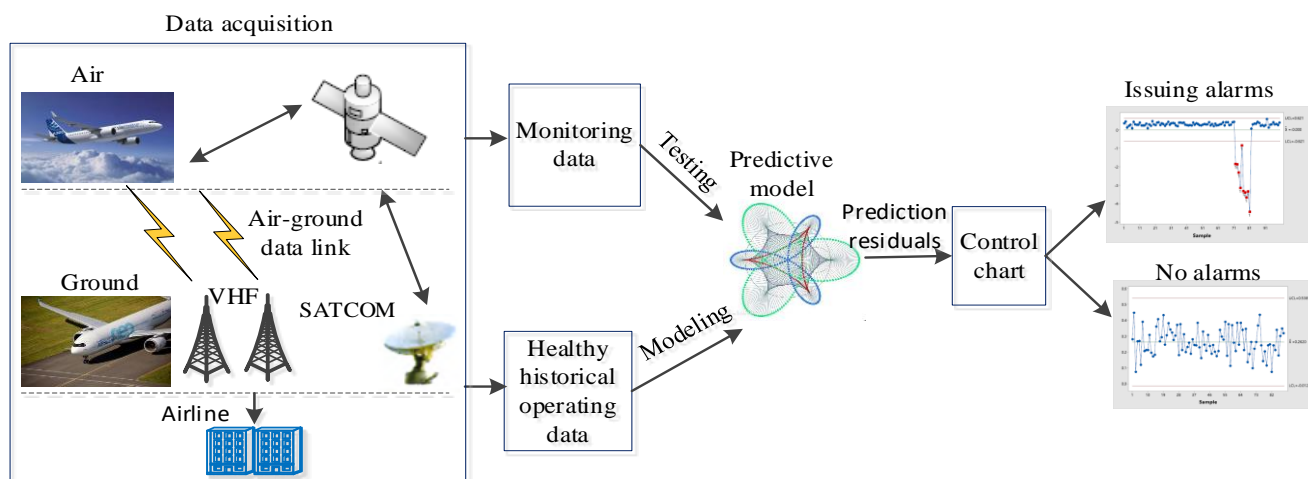

Fig. 1. The framework for fault prediction of BAS.

In Fig. 1, the BAS fault prediction framework consists of two components: the offline modeling process of the prediction model and the online warning process based on residuals. Aircraft operation data are transmitted back to the ground through the air-ground data link and obtained after being decoded by professional decoding software. Firstly, collected historical operational data are preprocessed, then the prediction models of the system feature variables are trained and the ones with the best prediction results are selected. In the online warning process, the EWMA control charts are used as the judging criteria for fault warning according to the prediction residuals of the prediction model after new operational data are obtained. If the prediction residuals exceed the limits, an alarm is generated to alert operating personnel to take measures to

ensure the safety of aircraft operations.

2.1. Prediction problem statements

To accurately determine the abnormal state of BAS in time, the state values of BAS feature variables are estimated based on normal historical operating data, and the following time series model is established:

$$\hat{y}_{T+1} = F(y_1, \ldots, y_T, x_1, \ldots, x_T) \tag{1}$$

where $\hat{y}_{T+1} \in \mathbb{R}$ denotes the predicted value of the target variable for the next moment. T is the size of the time window. y_t denotes the true value of the target variable at moment t. $Y = (y_1, \ldots, y_T) \in \mathbb{R}^T$ denotes the previous value of the target sequence within the time window T. $X = (x^1, \ldots, x^n)^T = (x_1, \ldots, x_T) \in \mathbb{R}^{n \times T}$ stands for a matrix of n relevant variables within the time window T. $x^k = (x_1^k, \ldots, x_T^k)^T \in \mathbb{R}^T$ denotes the

driving series of the kth variable within the time window T. $x_t = (x_t^1, \ldots, x_t^n)^T \in \mathbb{R}^n$ means a vector of n other relevant variables at moment t. x_t^k means the value of the kth variable at moment t. $F(\bullet)$ is a nonlinear mapping function, which takes the historical observation value Y of the target variable as well as other related variables X in the past as input to obtain the predicted value $\hat{y}_{T+1}$ of the target variable.

2.2. DSTP-ED model

The DSTP-ED model consists of an encoder-decoder and attention mechanisms, and both the encoder and decoder adopt LSTM. In the encoder, a two-phase attention mechanism is utilized to obtain the spatial association between the target and driving sequences. In the decoder, a temporal attention mechanism is employed to improve the response of the decoder to the long-term encoding vector.

2.2.1. Encoder with two-phase attention

Set the input sequence $X = (x_1, \ldots, x_T) \in \mathbb{R}^{n \times T}$, where n is the number of driving sequences, and T is the time window size. LSTM units are used for encoding to extract the feature expressions of the input sequence. Fig. 2 illustrates the structure of the encoder.
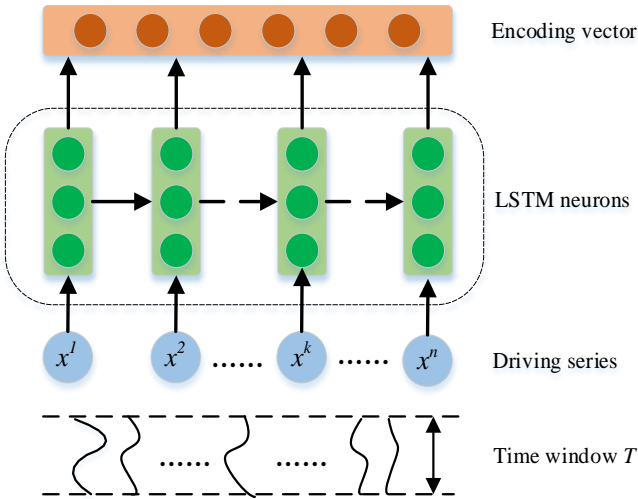

Fig. 2. The structure of the encoder.

In Fig. 2, the light green box is the LSTM neuron, while RNN and gated recurrent unit [6] can also be used as encoder neurons. The advantage of using LSTM neurons is that the gradient vanishing problem during the training of network parameters can be avoided and long-term dependencies can be better captured. Each LSTM unit has a hidden layer state h_t at time t, and there is also a memory state C_t for storing the hidden

layer state information of the LSTM unit before time t. The hidden layer states within the time window T together form the encoding vector. The inner architecture of the LSTM cell is described in Fig. 3.
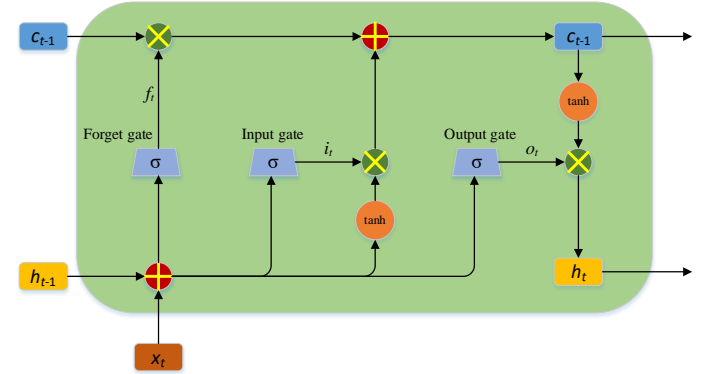

Fig. 3. The inner architecture of an LSTM cell.

The LSTM has a single storage unit and three gates to govern the information flow. Each LSTM cell receives the current moment's input x_t, the last moment's hidden layer state h_{t-1} and the last moment's memory state c_{t-1}. The following equation illustrates how the LSTM unit stores information and selectively lets it pass through.

$$f_t = \sigma(W_f[h_{t-1}; x_t] + b_f) \tag{2}$$

$$i_t = \sigma(W_i[h_{t-1}; x_t] + b_i) \tag{3}$$

$$o_t = \sigma(W_o[h_{t-1}; x_t] + b_o) \tag{4}$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_c[h_{t-1}; x_t] + b_c) \tag{5}$$

$$h_t = o_t \tanh(c_t) \tag{6}$$

Eq. (2) - (4) respectively represent the forget gate (f_t), the input gate (i_t) and the output gate (o_t) of the LSTM unit, respectively, which control the input and output of information. Eq. (5) - (6) describes how the cell state c_t and the currently hidden state h_t are updated. Where $[\; ; \;]$ is a concatenation operation. W_f, W_i, W_o, W_c , b_f, b_i, b_o, b_c are the trainable parameters, and σ is a sigmoid function.

(1) First phase attention module

This attention module belongs to the spatial attention mechanism used to obtain the spatial correlation between the driving sequences. Given the input driving sequence $x^k = (x_1^k, \ldots, x_T^k)^T \in \mathbb{R}^T$ of the kth variable, the softmax function is adopted to normalize the attention weights of all variables. The first phase attention mechanism is constructed as follows:

$$e_t^k = V_e^T \tanh(W_e[h_{t-1}^e; C_{t-1}^e] + U_e x^k + b_e), 1 \le k \le n \tag{7}$$

$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{j=1}^n \exp(e_t^j)} \tag{8}$$

where $h_{t-1}^e \in \mathbb{R}^m$ and $C_{t-1}^e \in \mathbb{R}^m$ are the previous hidden state and cell state of LSTM in the first phase attention mechanism of the encoder, respectively, and m is the size of the hidden state. $W_e \in \mathbb{R}^{T \times 2m}$, $U_e \in \mathbb{R}^{T \times T}$, V_e, $b_e \in \mathbb{R}^T$ are the network parameters that can be trained. e_t^k, α_t^k are the attention score and weight of the kth input driving sequence at time t, respectively. With the first phase attention weights, the input sequence is redefined as:

$$\tilde{x}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \ldots, \alpha_t^n x_t^n)^T \qquad (9)$$

Then, the hidden state h_t^e is updated to:

$$h_t^e = f_e(h_{t-1}^e, \tilde{x}_t) \qquad (10)$$

where f_e is the LSTM unit, h_t^e can be calculated according to Eq. (2) - (6), h_{t-1} needs to be replaced by h_{t-1}^e, x_t by $\tilde{x}_t$.

(2) Second phase attention module

This attention module also belongs to the spatial attention mechanism, which is used to obtain the spatial correlation between the driving and the target sequences. The output $\tilde{x}^k$ of the first phase attention mechanism is connected with the target sequences y^k at the same moment to obtain $z^k = [\tilde{x}^k; y^k] \in \mathbb{R}^{2 \times T}$, which is employed as an input to the second phase attention mechanism. The second phase attention weights are calculated as follows:

$$s_t^k = v_s^T \tanh(W_s[h_{t-1}^s; C_{t-1}^s] + U_s[\tilde{x}^k; y^k] + b_s), 1 \le k \le n \qquad (11)$$

$$\beta_t^k = \frac{exp(s_t^k)}{\sum_{j=1}^{n+1} exp(s_t^j)} \qquad (12)$$

where $h_{t-1}^s \in \mathbb{R}^p$ and $C_{t-1}^s \in \mathbb{R}^p$ are the previous hidden state and cell state of LSTM in the second phase attention module of the encoder, respectively, and p is the size of the hidden state of this attention module. s_t^k, β_t^k are the attention score and weight of the matrix z^k at time t, respectively. $W_s \in \mathbb{R}^{T \times 2p}$, $U_s \in \mathbb{R}^{T \times T}$ and v_s, $b_s \in \mathbb{R}^T$ are the trainable parameters. The output of the second phase attention mechanism is:

$$\tilde{z}_t = (\beta_t^1 z_t^1, \beta_t^2 z_t^2, \ldots, \beta_t^{n+1} z_t^{n+1})^T \qquad (13)$$

Then, the hidden state h_t^s is updated as:

$$h_t^s = f_s(h_{t-1}^s, \tilde{z}_t) \qquad (14)$$

where f_s is the LSTM unit, h_t^s can be calculated according to Eq. (2) - (6), h_{t-1} needs to be replaced by h_{t-1}^s, x_t by $\tilde{z}_t$.

2.2.2. Decoder with temporal attention

The decoder also adopts LSTM neurons, which use abstract features of the original input sequence obtained from the encoder to predict the output $\hat{y}_{T+1}$. A temporal attention mechanism is employed to improve the decoder's response to the long-time sequence encoding vectors by adaptively selecting the hidden states most relevant to the predicted target values. The specific temporal attention mechanism is calculated as the following:

$$l_t^i = v_d^T \tanh(W_d[d_{t-1}; c_{t-1}^o] + U_d h_i^s + b_d), 1 \le i \le T \qquad (15)$$

$$\gamma_t^i = \frac{\exp(l_t^i)}{\sum_{j=1}^T \exp(l_t^j)} \qquad (16)$$

where $d_{t-1} \in \mathbb{R}^q$ and $c_{t-1}^o \in \mathbb{R}^q$ are the previous hidden state and cell state of LSTM in the decoder, respectively, and q is the size of the hidden state in the temporal attention module. h_i^s is the hidden state of the ith encoder of the second phase attention module. l_t^i, γ_t^i are the attention score and weight of the hidden state h_i^s, respectively. The attention weight γ_t^i indicates the significance of the ith encoder hidden state for the prediction. V_d, $b_d \in \mathbb{R}^q$, $W_d \in \mathbb{R}^{p \times 2q}$ and $U_d \in \mathbb{R}^{q \times q}$ are the trainable parameters. Summing over the different encoder-weighted hidden states produces:

$$cn_t = \sum_{j=1}^T \gamma_t^j h_j^s \qquad (17)$$

Where $\gamma_t^j h_j^s$ denotes the weighted hidden state. cn_t is the context vector that represents the fusion information of hidden states of the encoder. And cn_t is different at each time step. Combining cn_t with the target series Y gives:

$$\tilde{y}_t = \tilde{w}^T[y_t; cn_t] + \tilde{b} \qquad (18)$$

where $\tilde{w} \in \mathbb{R}^{p+1}$ and $\tilde{b} \in \mathbb{R}$ are the parameters that map the concatenation to the decoder input. $\tilde{y}_{t-1}$ is employed to update the hidden state d_t of the decoder, as follows:

$$d_t = f_d(d_{t-1}, \tilde{y}_{t-1}) \qquad (19)$$

where f_d is the LSTM cell, d_t can be calculated according to Eq. (2) - (6), h_{t-1} needs to be replaced by d_{t-1}, x_t by $\tilde{y}_{t-1}$.

Finally, the context vector cn_t is connected to the decoder hidden state d_t as the new hidden state, and the linear function generates the prediction result:

$$\hat{y}_{T+1} = v_y^T(w_y[d_t; cn_t] + b_y) + b_y' \qquad (20)$$

The parameters $w_y \in \mathbb{R}^{q \times (q+p)}$ and $b_y \in \mathbb{R}^q$ map the concatenation to the size of the decoder hidden state, and the parameters $v_y \in \mathbb{R}^q$ and $b_y' \in \mathbb{R}$ are the weights and deviations of the linear function. The complete structure of the DSTP-ED model is depicted in Fig. 4. The parameters for training this encoder-decoder network are the learning process of the normal historical behavior. The decoder eventually predicts the next moment state of the target variable based on the relevant variables' historical and current information, as well as the historical data of the target variable.
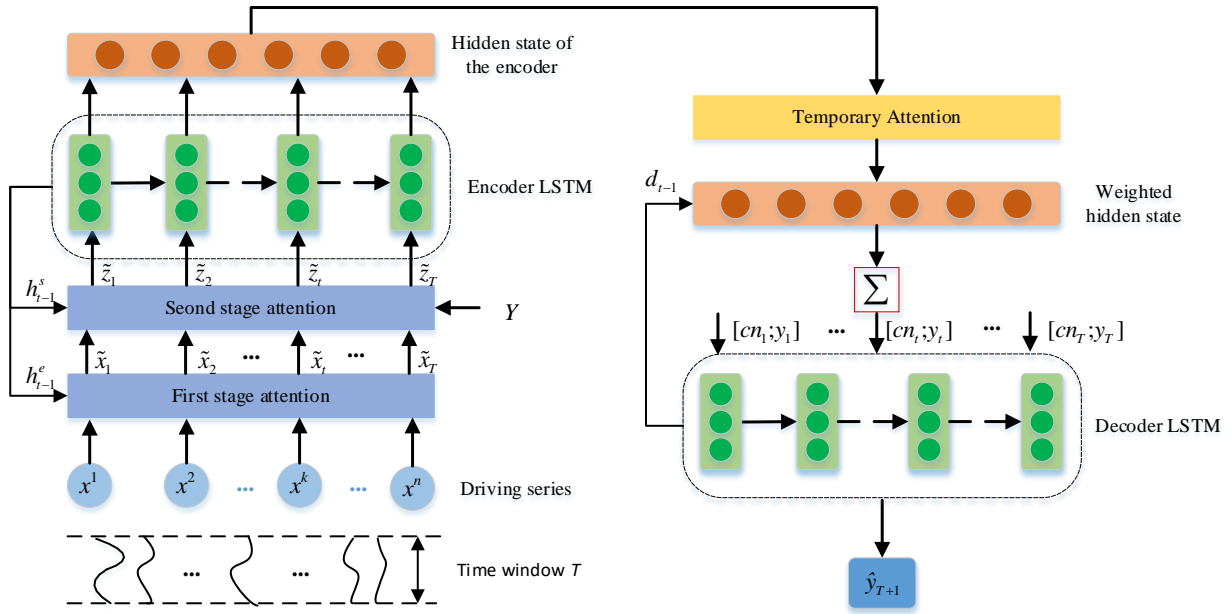
Fig. 4. Overall architecture of the DSTP-ED model.

2.3. Training procedure

The proposed DSTP-ED model is smooth and differentiable, so the standard backpropagation algorithm is adopted to train the model. The learning rate is a significant hyperparameter to be tuned while training the model. A novel cyclic learning rate is used in the training process, allowing the learning rate to vary cyclically between reasonable boundary values instead of decreasing monotonically [29]. Training with a cyclic learning rate rather than a constant value improves training accuracy without tuning and with fewer iterations. The batch size is set as 128. The minibatch stochastic gradient descent (SGD) [16] is employed to minimize the mean squared error (MSE) between the true value of the target variable y_t and the predicted value $\hat{y}_t$:

$$\vartheta(y_t, \hat{y}_t) = \frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2 \tag{21}$$

where N is the sample size.

2.4. Residuals-based monitoring method

The normal historical operational data from BAS are utilized for training the DSTP-ED model. When the health status of BAS changes, it will lead to an increase in the prediction error of the target variable, so such patterns can be identified by the residuals between the predicted and actual values, which is important for the prevention of BAS failures. The EWMA control chart is employed to draw criteria for detecting BAS

anomalies to eliminate the influence of operating conditions, sensor noise, etc., on the fault prediction results. The EWMA assigns different weights to the residual values at different moments to obtain the influence of the data at different moments on the current state. The closer the data are to the current moment, the larger the weighting factor. In addition, the EWMA is sensitive to small offsets and can quickly detect anomalies.

The statistic $z(t)$ of the EWMA control chart is expressed as:

$$z(t) = \lambda \Delta y_t + (1 - \lambda)z(t - 1) \tag{22}$$

where Δy_t is the predicted residual of the DSTP-ED model at time t. $\lambda \in (0,1)$ is the smoothing parameter. The smaller the λ, the smaller the bias can be detected, λ is generally set at 0.2-0.3 [20].

The initial value and standard deviation of the variable $z(t)$ are:

$$z_0 = \mu_0, \quad \delta_{z(t)} = \delta \sqrt{\frac{\lambda}{(2-\lambda)N} [1 - (1-\lambda)^{2t}]} \tag{23}$$

where μ_0, δ are the mean and standard deviation of the normal data prediction residuals, respectively, and N is the sample size. When $z(t)$ exceeds the control limits, an anomaly is considered to occur. The central and upper/lower control limits (CL, UCL, LCL) of the EWMA control chart are shown below:

$$CL(t) = \mu_0 \tag{24}$$

$$UCL(t)/LCL(t) = \mu_0 \pm L\delta \sqrt{\frac{\lambda}{(2-\lambda)N} [1 - (1-\lambda)^{2t}]} \tag{25}$$

where L determines the control limit range, which is usually set

to 3 [20]. The anomaly detection effectiveness of the EWMA control chart depends on the standard deviation δ, L, and λ. If the process mean is not drifting too fast, EWMA with suitable λ would be a good ahead predictor.

2.5. Fault prediction process

The complete fault prediction process includes three modules: data preparation, prediction model training and online warning, as shown in Fig. 5.
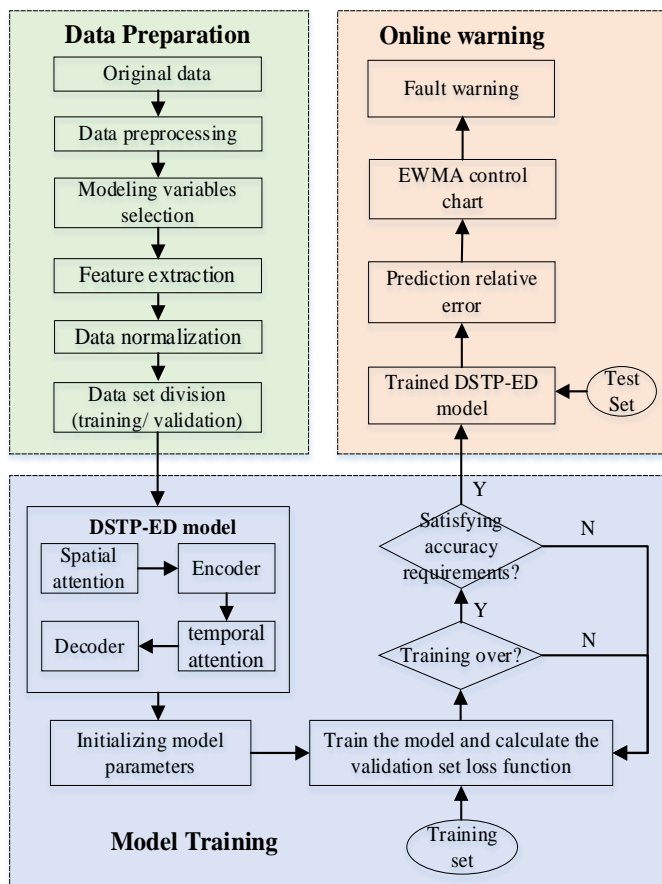

Fig. 5. BAS fault prediction process.

- Data preparation

Health data from BAS are collected and preprocessed in the data preparation phase, mainly involving missing value additions (due to different sensor sampling frequencies). Based on systematic principles and engineering experience, the modeling variables are screened and features are extracted to improve modeling efficiency. Since different parameters of BAS have different magnitudes, a normalization algorithm is used to transform the eigenvalues of BAS to between 0 and 1 to eliminate the effect of magnitudes. Finally, the processed data are divided into training and validation sets for input into the prediction model training module.

- Prediction model training

For the model training phase, the DSTP-ED model consists of a spatial attention module, an encoder, a temporal attention module and a decoder. After the training data are input to the model, the model parameters are initialized, the model is trained using the backward propagation algorithm, and the loss function of the validation set is calculated. When the training times reach the set Epoch, judge whether the loss function of the validation set reaches the accuracy requirement, if it does, the training is completed, otherwise, the training continues.

- Online warning

In the online warning stage, the test set data (new run data) are processed and input to the trained DSTP-ED model to obtain the prediction error of target variables. The EWMA control chart is employed to determine whether the prediction error exceeds the control limit and to alert flights that exceed the limit. Thus, BAS abnormalities are detected in advance and fault prediction is realized.

3. Application examples and result Analysis

3.1. Case study description

This study selects the BAS of the Airbus 320 series aircraft as an application case. A schematic of the BAS is depicted in Fig. 6.
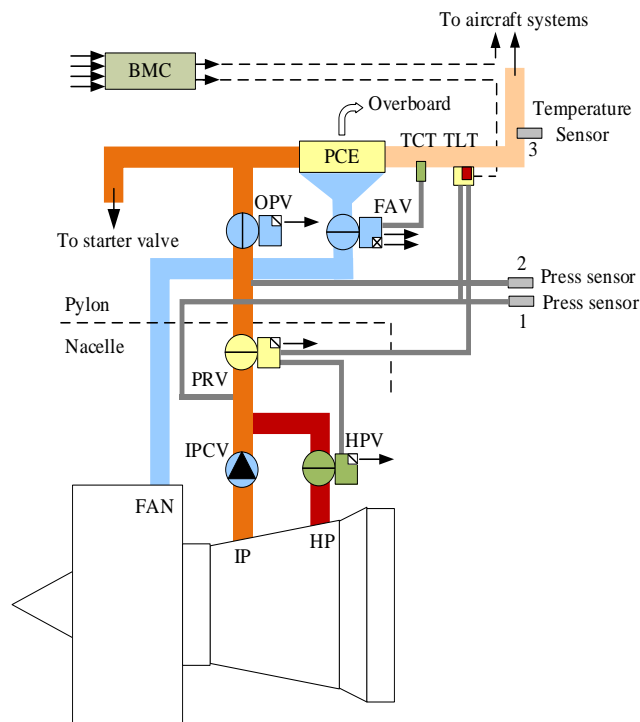

Fig. 6. Schematic drawing of the BAS.

Note: No. 1 and No. 2 are pressure sensors, and No. 3 is a temperature sensor. No. 1 sensor monitors the upstream of the PRV (pressure downstream of the HPV) and determines the HPV operating status. No. 2 sensor monitors the pressure downstream of the PRV. No. 3 monitors the temperature downstream of the PCE. HP stands for high pressure and IP stands for intermediate pressure.

For pressure regulation, the control components include the high-pressure bleed valve (HPV), intermediate-pressure bleed check valve (IPCV), over-pressure valve (OPV), and pressure regulator bleed valve (PRV) commanded by the temperature limitation thermostat (TLT). For temperature regulation, the control components include the pre-cooler exchanger (PCE), and fan air valve (FAV) commanded by the temperature control thermostat (TCT). The high temperature and pressure air from the engine compressor is reduced to a constant pressure of about 40 PSI and a constant temperature of about 200 degrees Celsius by the pressure and temperature regulator in the BAS, which is sent to the air intake manifold to provide to the downstream customers. The BAS is controlled by two bleed monitoring computers (BMCs) and the related information is sent to the central fault display system.

3.2. Data description and preprocessing

The experimental data are the real monitoring data (QAR data) generated during the operation of an Airbus A320 series aircraft. QAR is an airborne flight data recorder that can continuously record up to 600 hours of raw flight information and collect hundreds or even thousands of different flight parameters at the same time. The data recorded by QAR covers aircraft flight parameters (altitude, speed, acceleration, etc.), environmental parameters (temperature, air pressure, wind speed, etc.) and system performance parameters, etc. The number of QAR parameters exceeds three thousand. QAR data is the most comprehensive data recorded on aircraft today and has been widely used for flight quality analysis and aircraft equipment condition monitoring.

Each aircraft has two identical BASs, namely, the left BAS and the right BAS. The data sampling frequency for a single flight is 1 Hz, and the sampling time is the entire flight cycle. Historical data cover all operational states of BASs between November 2018 and October 2019, and its maintenance records

are presented in Table 1.

Table 1. Description of maintenance records.

Aircraft type	Maintenance records		Period	Total number of flights
A320-214	Left BAS	July 8th, 2019	Replacement of TCT	November 1st, 2018, to October 31st, 2019	1585
		August 4th, 2019	Replacement of PRV		
	Right BAS		No records		

Note: The maintenance record excludes general maintenance work, but only refers to the maintenance record after the failure.

Since QAR data are measured by sensors of airborne systems and the sampling frequency varies from sensor to sensor, some missing values exist in some parameters of the data, which need to be supplemented. According to the QAR data characteristics of BAS, this study uses linear interpolation to complement the data:

$$L(t) = y_u + \frac{y_v - y_u}{v - u}(t - u) \qquad (26)$$

where t is the point of missing data, u and v are the points without missing data before and after t, respectively; y_u and y_v are the values of u and v points; $L(t)$ is the result after interpolation.

3.3. Modeling variable selection

The main characteristic variables of BAS include bleed air pressure and bleed air temperature, and the influence brought by the external environment should also be considered. Based on the system operation mechanism and engineering experience, the following parameters are chosen as the feature variables of BAS, as shown in Table 2.

Table 2. List of BAS feature variables.

Variable attributes	Parameters	Description	Unit
Related variables	ALT	The flying altitude of the aircraft	ft
	N1	Engine low-pressure rotor speed	% rpm
	N2	Engine high-pressure rotor speed	% rpm
	SAT	Static air temperature	°C
	TAT	Total air temperature	°C
	MACH	Flight Mach number	/
Target variables	BAT	Bleed air temperature	psi
	BAP	Bleed air pressure	°C

3.4. Feature extraction

Fig. 7 shows the variation of two feature variables, BAT and BAP, over time for a single flight. BAT and BAP have different variation characteristics at different flight phases (other

parameters are similar). Therefore, BAS needs to be modeled in stages to improve the predictive performance of the DSTP-ED model. The flight phases are divided into four phases: ground, takeoff, cruise, and descent, depending on the flight altitude, engine speed, and Mach number. In this paper, QAR data are selected when the aircraft is on the cruise and flying at an altitude higher than 20,000 feet, the anti-icing system is off, and the left and right air conditioning components are on.
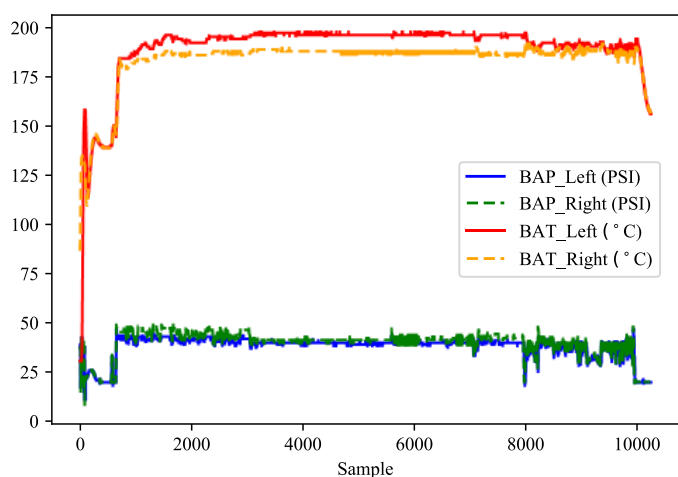

Fig. 7. Variation of BAT and BAP over time in a single flight.

Data of each flight cruise phase are sampled at intervals, and each feature variable of BAS is sampled continuously for 10s, and a total of 10 acquisitions are made at one-minute intervals each time. Data of each parameter continuously sampled for 10s are taken as the mean value as the eigenvalue data, as shown in Fig. 8 for BAT and BAP eigenvalue data of an aircraft in the 4-month cruise phase.
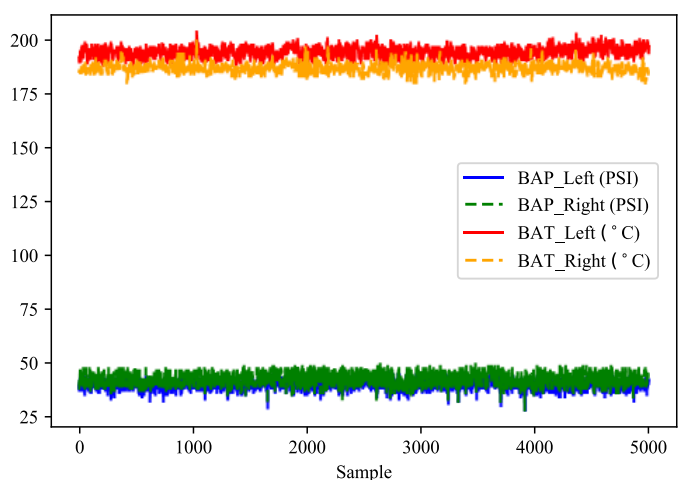

Fig. 8. Eigenvalue data of BAT and BAP for multiple flights in the cruise phase.

3.5. DSTP-ED model validation

Since the difference in the order of magnitude of the original data will introduce errors to the prediction model, which will easily cause the neurons to saturate and decrease network expression, the eigenvariables are normalized and reduced to [0,1]. The normalized equation is:

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{27}$$

where x is the original value of the sample data, x_{max} and x_{min} are the maximum and minimum values of the sample data, respectively, and x^* is the value after normalization.

The normalized eigenvalue data of all flight cruise phases in Table 1 are classified as the training set, validation set, and test set. The healthy data from November 2018 to June 2019 are used for the training set with 1038 flights; the data from September to October 2019 are taken as the validation set with 268 flights; the data containing faults from July to August 2019 are utilized for the test set with 279 flights.

There are four key parameters in the DSTP-ED model, including the time window size (T), the size of the hidden state of the encoder (m and p), and the size of the hidden state of the decoder (q). To simplify the complexity of the model, let $m = p = q \in \{16,32,64,128,256\}$, $T \in \{10,20,\dots,100\}$ and use grid search to find the parameters with the best model performance on the validation set. When $m=p=q$ is constant, the prediction results of BAP show the best performance at T=50, and the prediction results of BAT perform best at T=40, which is illustrated in Fig. 9. It can be seen that the model performance is worse when the time window is too long or too short. BAP and BAT are best predicted when T is fixed at $m = p = q = 16$, as described in Fig. 10.
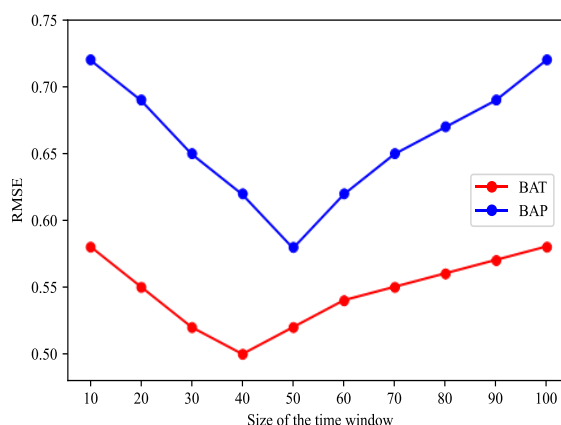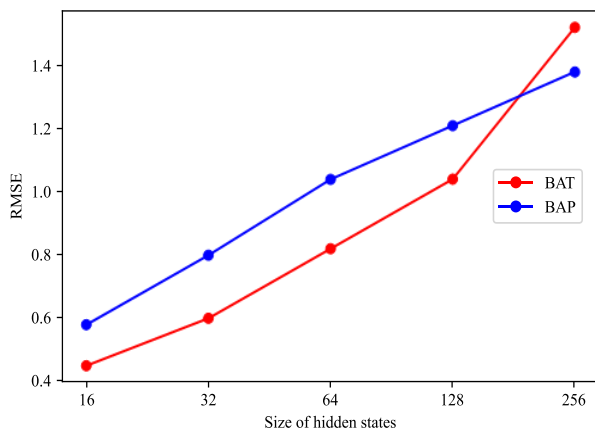

Fig. 9. RMSE vs. size of the time window.

Fig. 10. RMSE vs. size of hidden states.

Experiments are performed on data from the right BAS of the test set and compared with four other baseline algorithms (ARIMA, SVR, LSTM, ED) to verify the validity of the DSTP-ED model. ARIMA, in determining the model parameters, first requires differencing the non-stationary time series and converting them to a stationary series. Then, the autocorrelation and partial autocorrelation functions of the differenced series are analyzed to determine the order of autoregression and moving average [2]. The SVR, LSTM, and ED models use a grid search algorithm to determine the model parameters. The parameters mainly adjusted for SVR are "Penalty factor" and "Kernel parameter"; the parameters mainly adjusted for LSTM are "Number of neurons" and "Time window size"; the parameters mainly adjusted for ED are "Size of the hidden state of the encoder/decoder" and "Time window size". The parameter settings of each model are shown in Table 3 -Table 7.

Table 3. ARIMA model parameter settings.

Model parameters	BAP	BAT
Autoregressive order	1	2
Difference order	1	1
Moving average order	4	3

Table 4. SVR model parameter settings.

Model parameters	BAP	BAT
Penalty factor	5	10
Insensitive loss factor	0	0
Kernel function	Radial basis function	Radial basis function
Kernel parameter	1	1

Table 5. LSTM model parameter settings.

Model parameters	BAP	BAT
LSTM layers	1	1
Learning rate	0.01	0.01
Number of neurons	256	256
Time window size	5	5
Epochs	100	100
Batch size	128	128
Loss function	MSE	MSE
Optimization algorithm	Adaptive moment estimation	Adaptive moment estimation

Table 6. ED model parameter settings.

Model parameters	BAP	BAT
Size of the hidden state of the encoder	128	128
Size of the hidden state of the decoder	128	128
Time window size	10	10
Learning rate	0.01	0.01
Epochs	100	100
Batch size	128	128
Loss function	MSE	MSE
Optimization algorithm	Adaptive moment estimation	Adaptive moment estimation

Note: A single-layer LSTM is used for encoding and decoding the ED model.

Table 7. DSTP-ED model parameter settings.

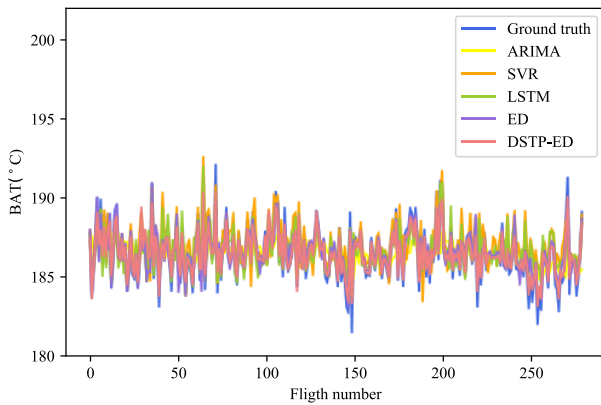| Model parameters	BAP	BAT
Size of the hidden state of the encoder	16	16
Size of the hidden state of the decoder	16	16
Time window size	50	40
Base learning rate	0.001	0.001
Maximum learning rate	0.1	0.1
Epochs	6000	6000
Batch size	128	128
Loss function	MSE	MSE
Optimization algorithm	SGD	SGD

Note: A single-layer LSTM is used for encoding and decoding the DSTP-ED model.

The prediction results and prediction errors of the DSTP-ED and the four baseline methods are shown in Fig. 11 and **Błąd! Nie można odnaleźć źródła odwołania.**, respectively. As Fig. 11 shows, the predicted values of each model have the same trend as the actual values and fluctuate up and down around the actual values. Although each model has a good prediction effect,

this does not reflect the real performance of the model. According to **Błąd! Nie można odnaleźć źródła odwołania.**, it can be seen that the prediction errors of each model are distributed around 0. Among them, ARIMA has the largest prediction error range, SVR and LSTM are the second, and ED and DSTP-ED have the smallest prediction error range. The above analysis indicates that ARIMA has a relatively poor prediction effect, SVR and LSTM have an intermediate prediction effect, and ED and DSTP-ED have a better prediction effect.
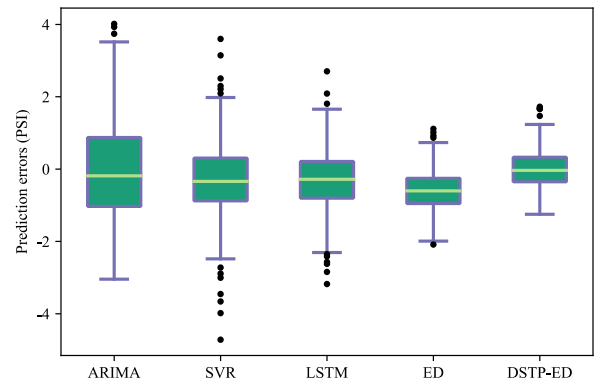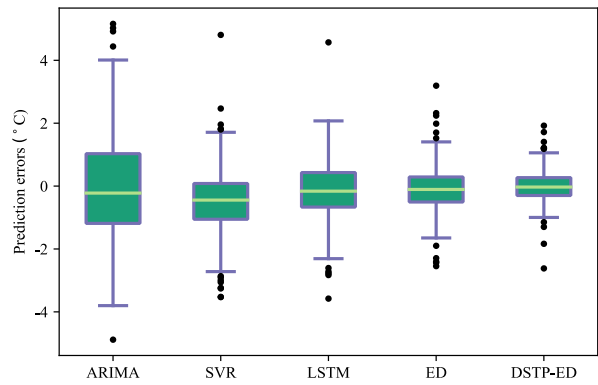

(a)   Prediction errors of BAP.
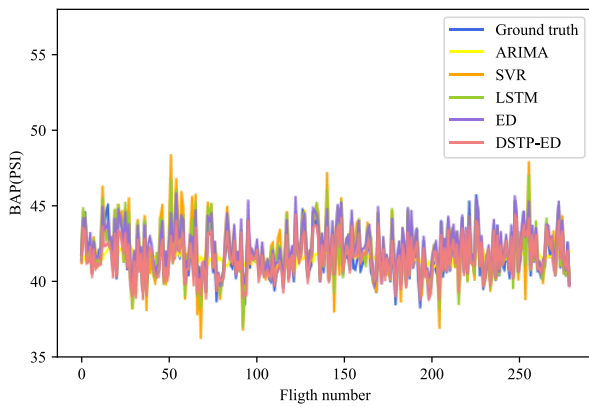
(a)   Prediction results of BAP.

(b)   Prediction errors of BAT.

Fig. 12. The prediction errors of different methods.

To further compare the performance of each model, the root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), as defined in Eqs. (28) - (30), are used as evaluation metrics for the models. The prediction performances of different methods are compared in Table 8.

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{y}_t - y_t)^2} \qquad (28)$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|\hat{y}_t - y_t| \qquad (29)$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \times 100\% \qquad (30)$$

where y_t is the true value of the target variable and $\hat{y}_t$ is the predicted value at time t.

(b)   Prediction results of BAT.

Fig. 11. The prediction results of different methods.

Table 8. Prediction performance using different methods of BAP and BAT.

Methods	BAP			BAT		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
ARIMA	1.15	2.74%	1.44	1.31	0.70%	1.67
SVR	0.85	2.03%	1.15	0.83	0.45%	1.10
LSTM	0.69	1.66%	0.91	0.70	0.37%	0.93

ED	0.67	1.62%	0.80	0.55	0.30%	0.75
DSTP-ED	0.44	1.05%	0.58	0.36	0.19%	0.50

Table 8 indicates that the DSTP-ED outperforms the other models with the optimal RMSE, MAE, and MAPE. Specifically, the ARIMA model performs much worse than any other model. This is mainly because the ARIMA model only focuses only on the target variable and ignores other variables, which does not apply to multivariate forecasting. The SVR model with the kernel-based approach focuses more on the spatial correlation between variables, while LSTM and ED models focus more on the time dependence, so the LSTM-based approach performs better. The ED model has a slightly better prediction effect than the LSTM model, but the improvement of the model is limited because the length of the context vector of the ED model is fixed and cannot handle too long time series. The DSTP-ED model integrates the dual-stage attention mechanisms in space and time and achieves better model performance compared with the ED model. This is because the employed spatial and temporal attention mechanisms can dynamically adjust the influence weights of the relevant variables on the target variables and properly capture the long time dependencies of the time series. The comparative analysis shows that the DSTP-ED model is more effective than other classical methods for predicting BAS data.
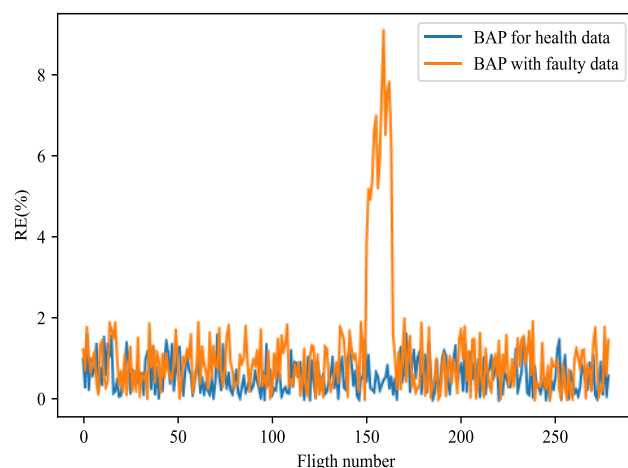
3.6.Monitoring results

The EWMA control chart is used to detect anomalies of the BAP and BAT prediction errors, which are measured by the relative error (RE). The expression of Δy_t in Eq. (22) is given by:

$$\Delta y_t = \text{RE} = \frac{|\hat{y}_t - y_t|}{y_t} \times 100\% \qquad (31)$$

where y_t is the true value and $\hat{y}_t$ is the predicted value at time t.

By comparing the RE of the DSTP-ED model for healthy BAS and faulty BAS, we can find that the RE of faulty BAS drifts before failure occurs, and **Błąd! Nie można odnaleźć źródła odwołania.** illustrates such a pattern. In **Błąd! Nie można odnaleźć źródła odwołania.**, the RE of BAP and BAT of abnormal data gradually deviates from the normal range, reaches a peak, and then returns to the normal range after maintenance, so the RE of the DSTP-ED model can be used to identify the potential BAS faults.
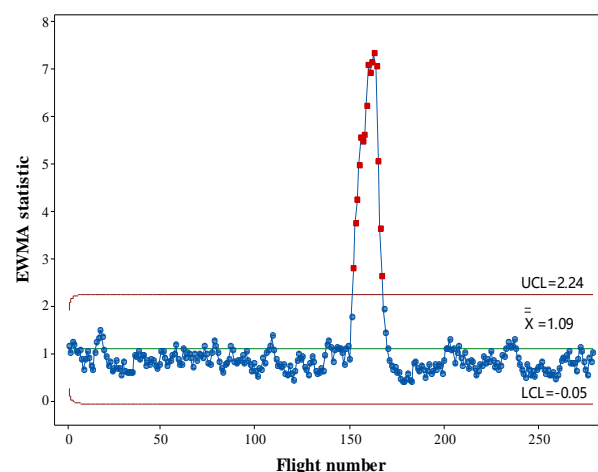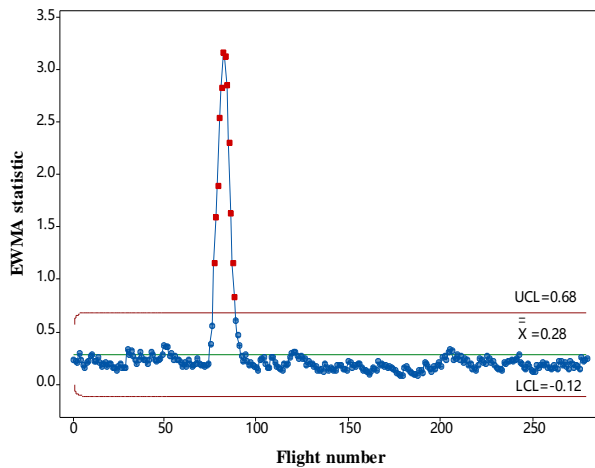

(a) REs of BAP.
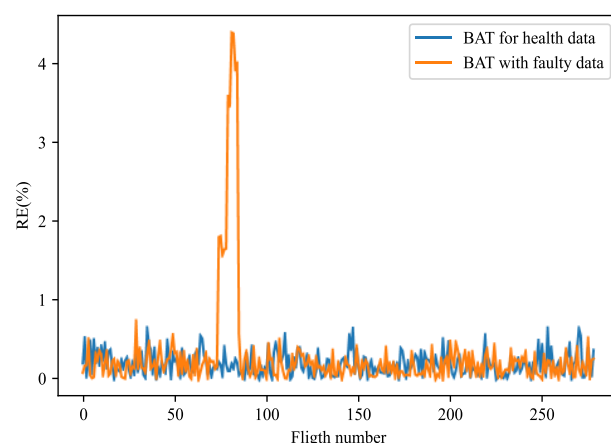
(b) REs of BAT.

Fig. 13. Comparison of REs of healthy BAS and BAS with faulty data.
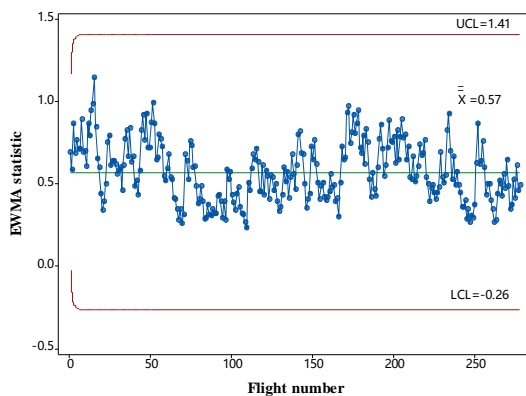
The RE of BAP and BAT in the test set data are monitored separately by EWMA control charts, and the monitoring results are illustrated in **Błąd! Nie można odnaleźć źródła odwołania.**.
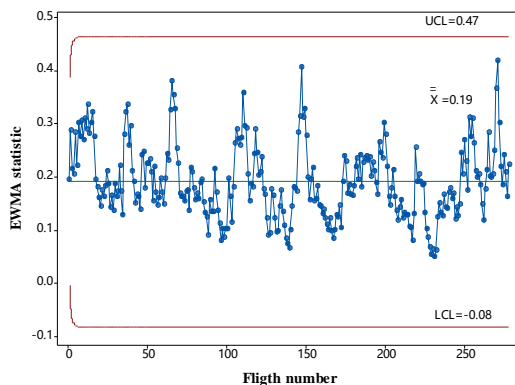
(a) EWMA control chart of BAP (left BAS).

(b) EWMA control chart of BAT (left BAS).

(c) EWMA control chart of BAP (right BAS).

(d) EWMA control chart of BAT (right BAS).
Fig. 14. EWMA control chart of BAS.

Błąd! Nie można odnaleźć źródła odwołania. (a) and (b) demonstrate the EWMA control chart for the left BAS, and **Błąd! Nie można odnaleźć źródła odwołania.** (c) and (d) describe the EWMA control charts for the right BAS. All EWMA statistics in Fig. 14 (c) and (d) are between UCL and LCL. This is consistent with the maintenance records, crew fault reports, and warning system records for the right BAS, with no false

alarms. As seen in **Błąd! Nie można odnaleźć źródła odwołania.** (a) and (b), flights 152 to 167 and flights 77 to 88 exceed the control limits. According to the maintenance records in Table 1, it can be seen that the maintenance was performed on July 8th and August 4th and the TCT and PRV were replaced, respectively. The maintenance on July 8th corresponds to the end of the 88th flight and the record on August 4th corresponds to the end of the 167th flight. Since operators can dispatch aircraft with faulty BAS as specified in the Minimum Equipment List, there is the possibility that the crew reported the corresponding failure before the maintenance was recorded.

We have reviewed the crew fault reports and the current system warnings and found that a BAS alert appeared in the warning system on flight 167 and a BAS anomaly was reported by the crew on flight 86. Therefore, combining the maintenance records, crew fault reports, and the current warning system, it can be concluded that the left BAS failures are all successfully identified a dozen flights (about two days) in advance. Additionally, according to Fig. 14, it can also be concluded that the repair on July 8th was caused by a BAT anomaly and the August 4th repair was due to a BAP anomaly, which can help the maintenance personnel locate the defective parts more quickly.
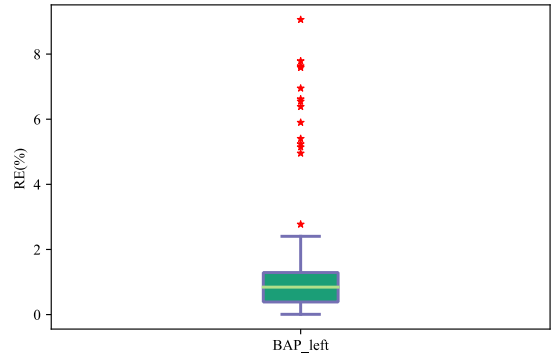
3.7. Comparison and discussion

To validate the effectiveness of the EWMA control chart for BAS anomaly detection, two other statistical methods, mean control chart (Xbar) and box plot, are selected for comparison. Xbar is the most common and basic control chart used to analyze the trend of changes in the center of the production process by determining whether the average value of the production process is in the required state of control. The box plot is a statistical method that shows the dispersion of data and is mainly used to reflect the characteristics of the original data distribution. It detects the data by five points: median, upper quartile, lower quartile, upper limit, and lower limit, and labels the information that exceeds the limits. The results of BAS anomaly detection for the Xbar control chart and box plot are given in **Błąd! Nie można odnaleźć źródła odwołania.** and **Błąd! Nie można odnaleźć źródła odwołania.**, respectively.
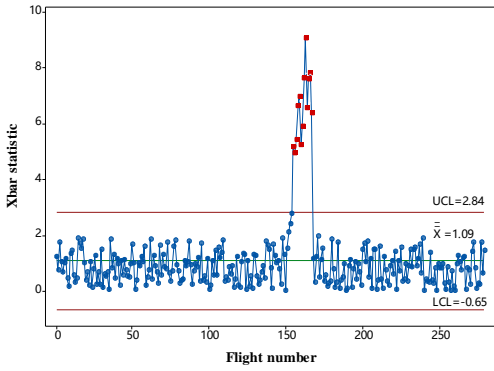
(a) Xbar control chart of BAP (left BAS).

(b) Xbar control chart of BAT (left BAS).

(c) Xbar control chart of BAP (right BAS).
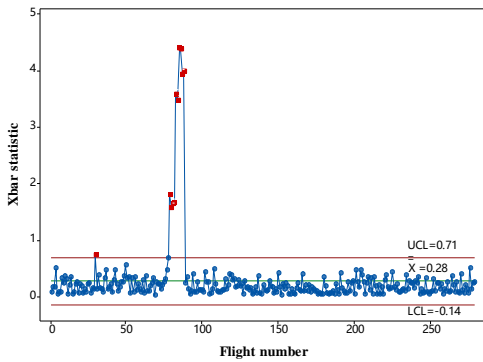

(d) Xbar control chart of BAT (right BAS).
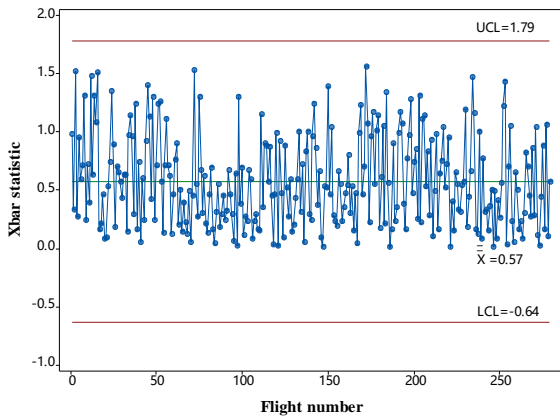
Fig. 15. Xbar control chart of BAS.

(a) Box plot of BAP (left BAS).
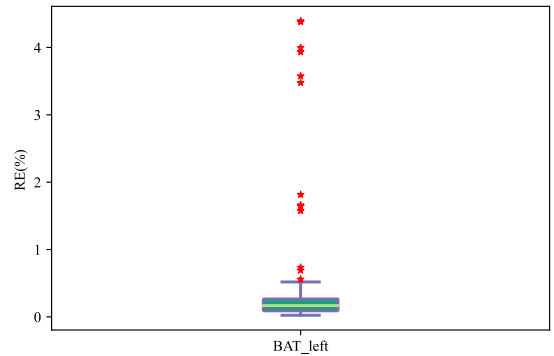
(b) Box plot of BAT (left BAS).
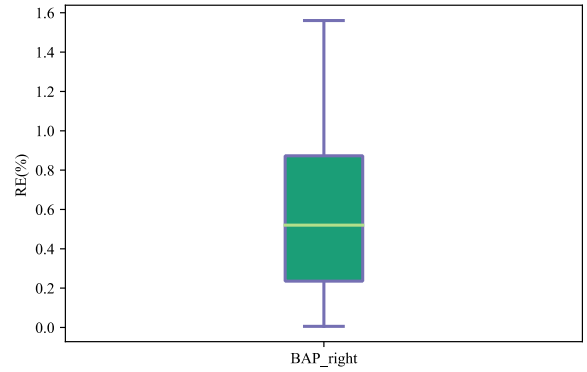
(c) Box plot of BAP (right BAS)
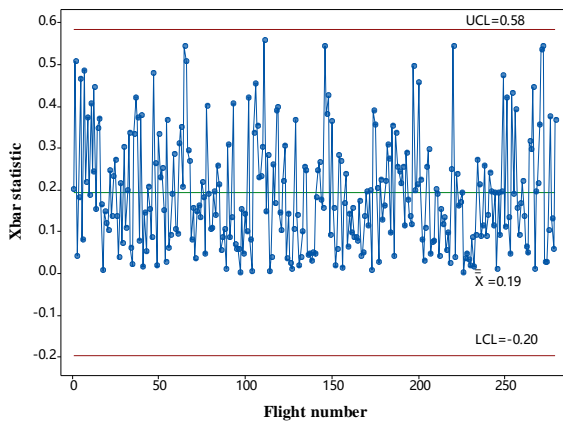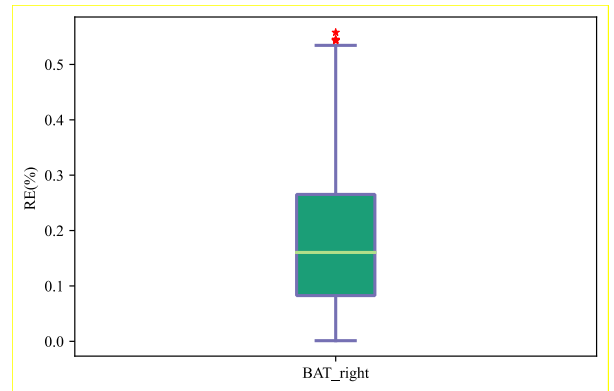
(d) Box plot of BAT (right BAS)

Fig. 16. Box plot of BAS.

In **Błąd! Nie można odnaleźć źródła odwołania.** (a),

flights 155-167 exceed the control limits. The Xbar control chart can detect BAS anomalies 12 flights earlier, referring to the analysis in Section 3.6. Compared to the EWMA control chart, Xbar's test results lagged by three flights. In **Błąd! Nie można odnaleźć źródła odwołania.** (b), flights 30, 79-88 exceed the control limits. The monitoring effect of Xbar lags behind EWMA by two flights, and there is a false alarm. As can be seen in **Błąd! Nie można odnaleźć źródła odwołania.** (c) and (d), the statistics of Xbar are within the control limits, which is consistent with the results of EWMA. **Błąd! Nie można odnaleźć źródła odwołania.** depicts the box plot of BAS. In **Błąd! Nie można odnaleźć źródła odwołania.** (a), 154-167 flight anomalies are detected, so the detection effect of the box plot is two flights later than EWMA and one flight earlier than Xbar. In **Błąd! Nie można odnaleźć źródła odwołania.** (b), flights 30, 50, 78-88 are detected as abnormal. The box plot can detect anomalies one flight earlier than Xbar, but it generates two false alarms. In **Błąd! Nie można odnaleźć źródła odwołania.** (c), no abnormal data are detected, and in **Błąd! Nie można odnaleźć źródła odwołania.** (d), two flights are detected as abnormal. However, the data in **Błąd! Nie można odnaleźć źródła odwołania.** (c) and (d) are all healthy, so the box plot produces two false detections. In comparison to the EWMA and Xbar control charts, the box plot has a high false alarm rate. The monitoring results of the three statistical methods are summarized in Table 9.

Table 9. Monitoring results of different methods.

Parameters		Methods	Alerts	False alarms
BAP	Left BAS	Box plot	Flights 154-167	0
		Xbar	Flights 155-167	0
		EWMA	Flights 152-167	0
	Right BAS	Box plot	None	0
		Xbar	None	0
		EWMA	None	0
BAT	Left BAS	Box plot	Flights 30, 50, 78-88	2
		Xbar	Flights 30, 79-88	1
		EWMA	Flights 77-88	0
	Right BAS	Box plot	Flights 111, 220	2
		Xbar	None	0
		EWMA	None	0

Based on the above analysis and Table 9, it is clear that the EWMA control chart has the best monitoring effect. Although all three monitoring methods detect faults in advance, the Xbar and box plot generate warning signals later than EWMA, and Xbar and box plot generate some error warnings. False alarms for Xbar and box plot may be caused by the operating environment or sensor noise, while EWMA does a good job of eliminating this effect and reducing false alarms. The comparison results indicate that the EWMA control chart is more advantageous in terms of warning time and accuracy and is more suitable to be used as a criterion for fault prediction. In conclusion, the proposed failure prediction method can successfully identify BAS faults before the failure occurrence and does not cause false alarms. Such early detection provides aircraft operators with sufficient time to take some measures. If the abnormal BAS can be maintained in time, the risk event can be prevented to ensure the safety of aircraft operation. The proposed method can also reduce some inefficient routine maintenance of BAS, reduce operating costs and improve operating efficiency.

Although the computational results support the effectiveness of the proposed method in predicting BAS failures, there are still some limitations in practical applications. Since the BAS operating state can be disturbed by the external environment and usage load, DSTP-ED prediction accuracy may be degraded, affecting fault warning effectiveness. However, the system state prediction model is not designed for real-time data flow, so the DSTP-ED model cannot update data online, which makes it less practical. Therefore, a prediction model with an online learning function can be considered in the future.

4. Conclusion

This paper proposed a BAS fault prediction method based on the DSTP-ED prediction model and EWMA control chart, which consists of two parts: offline modeling and online warning. In the offline modeling part, DSTP-ED was utilized to build an accurate state prediction model for BAS. In the online warning part, the prediction errors of BAS operation data were tracked by EWMA control charts to identify potential failures in advance effectively. The main conclusions were as follows:

(1) Taking the BAS of the Airbus 320 series as an example, real QAR data were selected to verify the prediction accuracy of the DSTP-ED model and the detection effect of the EWMA control chart. A total of eight modeling parameters were selected, with BAP and BAT as the key parameters reflecting the health status of BAS and the remaining parameters as auxiliary variables.

(2) The DSTP-ED model introduced the dual-stage two-phase attention mechanisms, improving the accuracy of the prediction results. The attention mechanisms could better capture the spatiotemporal relationships

between the driving and the target sequences, effectively solving the problems that the traditional model treated the input features with equal weight and the inability to handle long-time sequences. In the comparative analysis with other prediction models, the DSTP-ED obtained more accurate predictions with the lowest MAE, MAPE, and RMSE. Therefore, it was more suitable for predicting the state of BAS.

(3) The EWMA control chart was constructed based on the prediction deviation of the DSTP-ED model. The statistical process control results showed that the BAT and BAP prediction residuals of the healthy BAS were within the monitoring criteria, and impending BAS failures could be detected by abnormal BAT and BAP parameters about two days in advance. Compared with two classical statistical methods, the Xbar control chart and box plot, the EWMA control chart detected BAS abnormalities earlier, with fewer false alarms, and could effectively monitor the status of BAS online. The computational results supported the effectiveness of the proposed method in predicting BAS failures.

(4) The following work will validate the proposed failure prediction method on additional aircraft operational data and extend it to other systems (e.g., air conditioning systems, auxiliary power units, etc.) to reduce unexpected flight interruptions and improve operational safety and efficiency. In addition, the fault prediction model with online adaptive function needs further research.

Nomenclature

Symbols

Symbol	Description	Symbol	Description
c_t	cell state of LSTM	Δy_t	the predicted residual of the DSTP-ED model at time t
cn_t	the context vector	$\hat{y}_{T+1}$	the predicted value of the target variable for the next moment
c_t^o, d_t	the cell state and the hidden state of LSTM in the decoder, respectively	y^k	the corresponding target sequence of x^k
e_t^k, α_t^k	the attention score and weight of the kth driving sequence at time t, respectively	$\tilde{y}_t$	the input of the decoder
f_t	forget gate	z^k	the concatenation vector of x^k and y^k (the input of second phase attention)
f_e, f_s, f_d	LSTM unit	z_t	the concatenation vector of x_t and y_t
h_t	hidden state	$\tilde{z}_t$	z_t with second-phase attention weights (the output of second-phase attention)
h_t^e	the hidden state of LSTM (first phase attention module)	$z(t)$	statistics of the EWMA control chart
h_t^s	the hidden state of LSTM (second phase attention module)	z_0	the initial value of $z(t)$
i_t	input gate	C_t^e	the cell state of LSTM (first phase attention module)
l_t^i, γ_t^i	the temporal attention score and weight of the hidden state h_i^s, respectively	C_t^s	the cell state of LSTM (second phase attention module)
m	the size of the hidden state (first phase attention module)	CL	central limit
o_t	output gate	$F(\cdot)$	nonlinear mapping function
p	the size of the hidden state (second phase attention module)	L	the parameter that determines the control limit range
q	the size of the hidden state (temporal attention module)	LCL	lower Control Limit
s_t^k, β_t^k	the attention score and weight of the matrix z^k at time t, respectively	$LI(t)$	the result after interpolation
v_y, b_y'	the weights and deviations of the linear function	N	the size of the training data
w_y, b_y	the parameters that map the concatenation $[d_t; cn_t]$ to the size of the decoder's hidden state	T	the time window size
$\tilde{w}, \tilde{b}$	the parameters that map the concatenation $[y_t; cn_t]$ to the decoder input	UCL	upper control limit
x_t	a vector of n other relevant variables at moment t	$W_c, W_f, W_i, W_o,$ $W_d, U_d, V_d, b_d,$ $W_e, U_e, V_e, b_e,$ $W_s, U_s, V_s, b_s,$ $b_c, b_f, b_i, b_o,$	network parameters
$\tilde{x}_t$	x_t with first-phase attention weights	X	a matrix of n relevant variables within the time window T

Symbol	Description	Symbol	Description	Symbol	Description
x^k	the driving series of the kth variable within the time window T (the input of first phase attention)	Y	the previous value of the target sequence within the time window size T	x	the original value of the sample data	σ	prediction residuals, respectively / sigmoid function
$\tilde{x}^k$	the weighted driving series of the kth variable within the time window T (the output of first phase attention)	λ	smoothing parameter	x_{max}, x_{min}	the maximum and minimum values of the sample data	$\vartheta(y_t, \hat{y}_t)$	mean square error of predicted and true values
				x^*	the value after normalization	$[\ ;\ ]$	concatenation operation
x_t^k	the value of the kth variable at moment t	μ_0, δ	the mean and standard deviation of the normal data	y_t	the true value of the target variable at moment t		

Abbreviations

ACARS	aircraft communications addressing and reporting system	MAPE	mean absolute percentage error
ALT	flying altitude of the aircraft	MSE	mean squared error
ARIMA	autoregressive integrated moving average	N1	engine low-pressure rotor speed
BAP	bleed air pressure	N2	engine high-pressure rotor speed
BAS	bleed air system	OPV	Over-pressure valve
BAT	bleed air temperature	PRV	pressure regulator bleed valve
BMC	bleed monitoring computer	PCE	pre-cooler exchanger
DSTP-ED	dual-stage two-phase attention-based encoder-decoder	QAR	quick access recorder
ED	encoder-decoder	RE	relative error
EWMA	exponentially weighted moving average	RMSE	root mean squared error
FAV	fan air valve	RNN	recurrent neural network
HP	high pressure	SAT	static air temperature
HPV	high-pressure bleed valve	SVR	support vector regression
IP	Intermediate pressure	SGD	stochastic gradient descent
IPCV	Intermediate-pressure bleed check valve	TAT	total air temperature
LSTM	long short-term memory	TLT	temperature limitation thermostat
MACH	flight Mach number	TCT	temperature control thermostat
MAE	mean absolute error	Xbar	mean control chart

Acknowledgments

Reference

1. Abdelrahman WG, Al-Garni AZ, Al-Wadiee W. Application of Back Propagation Neural Network Algorithms on Modeling Failure of B-737 Bleed Air System Valves in Desert Conditions. Applied Mechanics & Materials 2012; 225: 505-510, http://dx.doi.org/10.4028/www.scientific.net/AMM.225.505.

2. Asteriou D, Hall S. ARIMA Models and the Box-Jenkins Methodology. Applied Econometrics 2011, 2(2): 265-286, http://dx.doi.org/10.1057/978-1-137-41547-9_13.

3. Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. Computer Science 2014; 1409, https://arxiv.org/abs/1409.0473v7.

4. Cheng Y, Huang F, Zhou L, Jin C, Yuejie Z, Tao Z, editors. A hierarchical multimodal attention-based neural network for image captioning. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information; 2017; Retrieval, http://dx.doi.org/10.1145/3077136.3080671.

5. Cho K, Merrienboer Bv, Bahdanau D, Bengio Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. Computer Science 2014, https://doi.org/10.48550/arXiv.1409.1259.

6. Cho K, Merrienboer Bv, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. 2014 Conference on Empirical Methods in Natural Language Processing; 25-29 October, 2014; Doha, Qatar: Association for Computational Linguistics (ACL); 2014; 1724-1734, http://dx.doi.org/10.3115/v1/D14-1179.

7. Di Gangi M, Federico M. Deep Neural Machine Translation with Weakly-Recurrent Units. EAMT 2018, https://arXiv:1805.04185.

8. Dokas IM, Karras DA, Panagiotakopoulos DC. Fault tree analysis and fuzzy expert systems: Early warning and emergency response of landfill operations. Environmental Modelling & Software 2009; 24(1): 8-25, https://doi.org/10.1016/j.envsoft.2008.04.011.

9. Elman JL. Distributed representations, simple recurrent networks, and grammatical structure. Mach Learn 1991; 7(2): 195-225, https://doi.org/10.1007/BF00114844.

10. Feng L, Wang H, Si X, Zou H. A State-Space-Based Prognostic Model for Hidden and Age-Dependent Nonlinear Degradation Process. IEEE Transactions on Automation Science and Engineering 2013; 10(4): 1072-1086, http://dx.doi.org/10.1109/TASE.2012.2227960.

11. Graves A, Mohamed A, Hinton G, editors. Speech recognition with deep recurrent neural networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; 2013 26-31 May 2013, http://dx.doi.org/10.1109/ICASSP.2013.6638947.

12. Guo T, Lin T. Multi-variable LSTM neural network for autoregressive exogenous model. 2018, https://arxiv.org/abs/1806.06384.

13. Guo Y, Sun Y, He Y, Du F, Su S, Peng C. A Data-driven Integrated Safety Risk Warning Model based on Deep Learning for Civil Aircraft. IEEE Transactions on Aerospace and Electronic Systems 2022: 1-14, http://dx.doi.org/10.1109/TAES.2022.3204224.

14. Hübner R, Steinhauser M, Lehle C. A Dual-Stage Two-Phase Model of Selective Attention. Psychological review 2010; 117: 759-784, http://dx.doi.org/10.1037/a0019471.

15. Karpathy A, Fei-Fei L. Deep Visual-Semantic Alignments for Generating Image Descriptions. Ieee T Pattern Anal 2017; 39(4): 664-676, http://dx.doi.org/10.1109/TPAMI.2016.2598339.

16. Kingma DP, Ba JL, editors. Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015, May 7, 2015 - May 9, 2015; 2015; San Diego, CA, United states: International Conference on Learning Representations, ICLR, https://arxiv.org/abs/1412.6980v2.

17. Liang Y, Ke S, Zhang J, Yi X, Zheng Y, editors. GeoMAN: Multi-level Attention Networks for Geo-sensory Time Series Prediction. Twenty-Seventh International Joint Conference on Artificial Intelligence 2018, http://dx.doi.org/10.24963/ijcai.2018/476.

18. Liu J, Zio E. SVM hyperparameters tuning for recursive multi-step-ahead prediction. Neural Computing and Applications 2017; 28(12): 3749-3763, http://dx.doi.org/10.1007/s00521-016-2272-1.

19. Liu Y, Gong C, Yang L, Chen Y. DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. Expert Syst Appl 2020; 143, http://dx.doi.org/10.1016/j.eswa.2019.113082.

20. Montgomery DC. Introduction to Statistical Quality Control. NewYork: John Wiley & Sons; 2007.

21. Orchard M, Vachtsevanos G. A particle-filtering approach for on-line fault diagnosis and failure prognosis. Transactions of The Institute of Measurement and Control - TRANS INST MEASURE CONTROL 2009; 31, http://dx.doi.org/10.1177/0142331208092026.

22. Pawełczyk M, Fulara S, Sepe M, De Luca A, Bad ora M. Industrial gas turbine operating parameters monitoring and data-driven prediction. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (3): 391–399, http://dx.doi.org/10.17531/ein.2020.3.2.

23. Peltier V, Dullenkopf K, Bauer H-J, editors. Numerical Investigation of the Aerodynamic Behaviour of a Compressor Bleed-Air System. ASME Turbo Expo 2014: Turbine Technical Conference and Exposition; 2014 June 16–20, http://dx.doi.org/10.1115/gt2014-25822.

24. Peltier V, Dullenkopf K, Bauer HJ, editors. Experimental investigation of the performance of different bleed air system designs. ASME Turbo Expo 2012: Turbine Technical Conference and Exposition, GT 2012, June 11, 2012 - June 15, 2012; Copenhagen, Denmark: American Society of Mechanical Engineers (ASME) 2012, http://dx.doi.org/10.1115/GT2012-68242.

25. Qin Y, Song D, Chen H, Cheng W, Jiang G, Cottrell G. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. IJCAI2017, https://arxiv.org/abs/1704.02971v1. https://doi.org/10.24963/ijcai.2017/366

26. Ramirez IS, Mohammadi-Ivatloo B, Marquez FPG. Alarms management by supervisory control and data acquisition system for wind turbines. Eksploatacja i Niezawodnosc 2021; 23(1): 110-6, http://dx.doi.org/10.17531/EIN.2021.1.12.

27. Shang L, Liu G, editors. Heat exchanger fouling detection in a simulated aircraft engine bleed air temperature control system. 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics 2010, http://dx.doi.org/10.1109/AIM.2010.5695895.

28. Shang L, Liu G. Sensor and Actuator Fault Detection and Isolation for a High Performance Aircraft Engine Bleed Air Temperature Control

System. IEEE Transactions on Control Systems Technology 2011; 19(5): 1260-1268, http://dx.doi.org/10.1109/TCST.2010.2076353.

29. Smith LN, editor Cyclical learning rates for training neural networks. 17th IEEE Winter Conference on Applications of Computer Vision, WACV 2017, March 24, 2017 - March 31, 2017; 2017; Santa Rosa, CA, United states: Institute of Electrical and Electronics Engineers Inc., http://dx.doi.org/10.1109/WACV.2017.58.

30. Smith M, Sulcs P, Walthall R, Mosher M, Kacprzynski G. Design and Implementation of Aircraft System Health Management (ASHM) Utilizing Existing Data Feeds. SAE 2015 AeroTech Congress & Exhibition 2015, https://doi.org/10.4271/2015-01-2587.

31. Su S, Sun Y, Li L, Peng C, Zhang H, Zhang T. Risk Warning for Aircraft Bleed Air System with Multivariate State Estimation Technique. Journal of Aerospace Information Systems 2022; 19(8): 550-564, https://doi.org/10.2514/1.I011011.

32. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. Advances in neural information processing systems 2014; 27.

33. Swanson DC. A general prognostic tracking algorithm for predictive maintenance. 2001 IEEE Aerospace Conference Proceedings (Cat No01TH8542) 2001; 6: 2971-2977, https://doi.org/10.1109/AERO.2001.931317.

34. Wang H, Peng M-j, Wesley Hines J, Zheng G-y, Liu Y-k, Upadhyaya BR. A hybrid fault diagnosis methodology with support vector machine and improved particle swarm optimization for nuclear power plants. Isa T 2019; 95: 358-371, https://doi.org/10.1016/j.isatra.2019.05.016.

35. Wang WQ, Golnaraghi MF, Ismail F. Prognosis of machine health condition using neuro-fuzzy systems. Mech Syst Signal Pr 2004; 18(4): 813-831, https://doi.org/10.1016/S0888-3270(03)00079-7.

36. Yan J, Koç M, Lee J. A prognostic algorithm for machine performance assessment and its application. Production Planning & Control - PRODUCTION PLANNING CONTROL 2004; 15: 796-801, https://doi.org/10.1080/09537280412331309208.

37. Yang B-S, Widodo A. Support Vector Machine for Machine Fault Diagnosis and Prognosis. Journal of System Design and Dynamics 2008; 2: 12-23, https://doi.org/10.1299/jsdd.2.12.

38. Yang C, Lou Q, Liu J, Yang Y, Cheng Q. Particle filtering-based methods for time to failure estimation with a real-world prognostic application. Applied Intelligence 2018; 48(8): 2516-2526, https://doi.org/10.1007/s10489-017-1083-0.

39. Yao R, Guo C, Deng W, Zhao H. A novel mathematical morphology spectrum entropy based on scale-adaptive techniques. Isa T 2022; 126: 691-702, https://doi.org/10.1016/j.isatra.2021.07.017.

40. Zhang Z, Verma A, Kusiak A. Fault Analysis and Condition Monitoring of the Wind Turbine Gearbox. IEEE Transactions on Energy Conversion 2012; 27(2): 526-535, https://doi.org/10.1109/TEC.2012.2189887.

41. Zhao H, Liu J, Chen H, Chen J, Li Y, Xu J, et al. Intelligent Diagnosis Using Continuous Wavelet Transform and Gauss Convolutional Deep Belief Network. IEEE Transactions on Reliability 2022; 1-11, https://doi.org/10.1109/TR.2022.3180273.

42. Zuber N, Bajri R. Gearbox faults feature selection and severity classificatiousing machine learning. Eksploatacja i Niezawodnosc 2020; 22(4): 748-56, https://doi.org/10.17531/ein.2020.4.19.