ISBN number 978-1-925627-90-9

Please select category below:
Normal Paper ☐
Student Paper ☒
Young Engineer Paper ☐

A Spatiotemporal Data Fusion Technique for Aircraft Environmental and Operational Condition (EOC) Representation

Joelle W.Y. Chia, Wim J. C. Verhagen, Jose M. Silva and Ivan S. Cole

RMIT University, Melbourne, Victoria, 3000, Australia

Abstract

An aircraft undergoes both operational and non-operational phases throughout its lifecycle. During operational phases (i.e., when the aircraft is flying or the engine is running), operational flight data are recorded in modern cockpit instruments and monitoring systems. Conversely, limited information is captured during non-operational phases, resulting in an incomplete understanding of the Environmental and Operational Conditions (EOC) an aircraft is exposed to throughout its lifecycle. This paper addresses this challenge by proposing a data-driven approach to pre-process and fuse the spatiotemporal flight and weather data from different sources to construct a representative profile of the aircraft's EOC in a multivariate time series. The proposed framework employs a temporal resampling method and spatial alignment using the nearest neighbour method. Applying the framework using a Cessna 172S single-engine flight training aircraft's flight data and operational profile, the initially combined data was incomplete due to partially missing parameters. Consequently, a further processing step involves estimating and imputing the missing parameters from the available data. In this work, temperature estimation was performed using the linear regression model and the expectation-maximization algorithm. By capturing the continuous sequences of environmental and operational parameters, the combined and processed data offers a comprehensive understanding of operational usage severity and aircraft microclimates. This foundational work paves the way for future research in predicting aircraft system health by integrating environmental and operational variations for enhanced prognostics and health management.

Keywords: Aircraft environmental and operational conditions, Multivariate time-series, PHM, Predictive Maintenance, Spatiotemporal data fusion, Statistical Learning.

Introduction

The exposure of an aircraft to varying environmental and operational conditions (EOC) during its lifecycle is critical for evaluating its performance and health. Military aircraft typically employ Operational Load Monitoring (OLM) and Structural Health Monitoring (SHM) sensors for this reason. A prime example is the Joint Strike Fighter F-35 program [1], where the fighter jet is widely instrumented to collect environmental and operational data for Prognostics and Health Management (PHM). However, the cost analysis for installing SHM sensors could be suboptimal [2] depending on the type of aircraft and mission profiles, thereby limiting the applicability of this approach especially for civil and general aviation. In modern aircraft, flight data are collected in various aircraft systems. Thus, this limitation can be circumvented by analysing flight parameters from existing sources such as the flight data recorder [3], [4], quick access recorder [5], digital avionics instrument [6], [7] or communication systems [8], and ADS-B data [9]. In these previous works, data-driven approaches, including supervised learning, unsupervised learning, and deep learning models, have been applied to analyse the aircraft's operational conditions using flight data for anomaly detection and PHM.

While operational data are often recorded during flights, limited information is available for non-operational phases. This neglects the environmental impacts experienced during non-operational phases, such as when aircraft are parked on the ground or undergoing maintenance. Structural and component degradation caused by factors such as corrosion, ultraviolet exposure, or thermal cycling can occur during these periods, as evidenced by the widespread corrosion in grounded fleets during the COVID-19 pandemic [10]. Moreover, Digital Twins are becoming increasingly explored for aircraft maintenance and operations optimisation through PHM and Integrated Systems Health Management (ISHM) [11], [12], [13], [14]. The combination of both operational and non-operational data relating to the aircraft's environmental and operational conditions can help shape a more accurate digital representation of the physical aircraft's health state. System-wise, Wang et al. [15] evaluated environmental factors collected from airport bases to predict the remaining useful life of pitot tubes, which are prone to damage due to environmental impact. For aircraft structures, Trueman et al. [16] proposed a Corrosion Prognostics and Health Management (CPHM) system that enables continuous monitoring using environmental sensors and witness plates installed in enclosed areas within the aircraft and in their respective ground stations. This approach ensures a holistic coverage of environmental conditions that can contribute to significant long-term structural damages throughout the aircraft's service life. Nevertheless, such a system presents inherent logistical, certification, and technical challenges to be deployed in the field, especially in civil aircraft.

This paper aims to address the gaps highlighted by proposing a novel framework that (i) constructs a representative Environmental and Operational Conditions (EOC) profile that accounts for both operational and non-operational phases of an aircraft's lifecycle, (ii) without additional monitoring sensors, (iii) by focusing on data pre-processing. This framework is demonstrated in the context of analysing the EOC that affects structural corrosion on a Cessna 172S general aviation aircraft.

Data Fusion Framework and Methodology

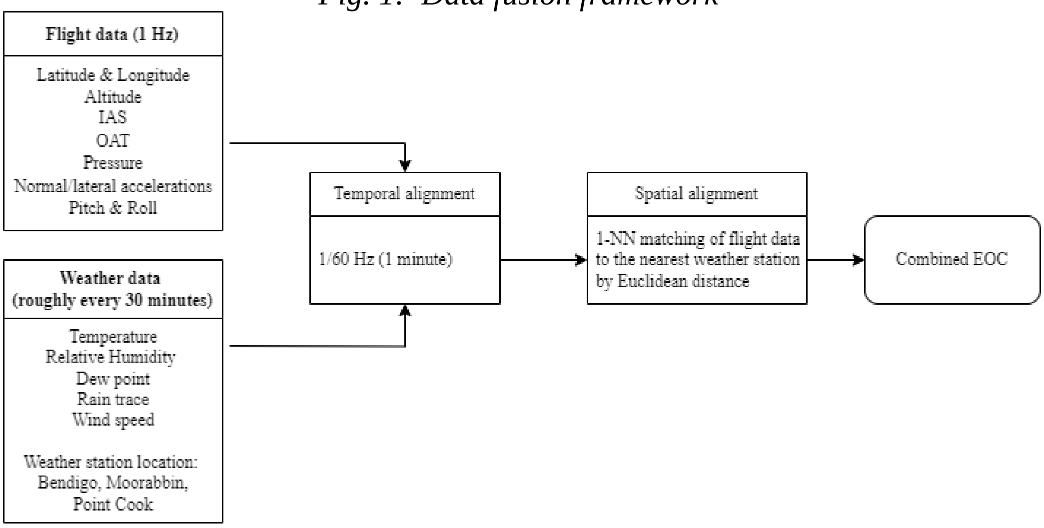
In this paper, the flight data source comes from the Cessna 172S Garmin G1000 system. The G1000 system's main function is to serve as the electronic flight instrument system, providing pilots with real-time information on the flight path, the aircraft's systems, and the environment. Nevertheless, the system also maintains logs of each engine start which can be downloaded from the memory card. The recorded data acquisition rate is 1 Hz, i.e. data is logged every second. For more background on the nature of these data, Fala et al. [17] described the G1000 dataset as they applied machine learning models to these data for unsupervised flight phase identification.

Relevant environmental variables are obtained from local weather stations of the aircraft's bases and frequent flight destinations. In this paper, the authors collected weather information from the Australian Bureau of Meteorology's website [18] from three selected weather stations: Bendigo, Moorabbin, and Point Cook in Victoria, Australia. The weather data are updated roughly every 30 minutes, but the frequency could vary.

Fig. 1 illustrates the data fusion pre-processing framework with the parameters extracted from each data source and their respective frequencies. Since both data sources are represented as multivariate time-series, the first step would be temporal alignment of the two datasets. Giving more importance to operational variation, a common standardised frequency is set at $\frac{1}{60}$ Hz (i.e. 60 seconds). Time-series resampling is a common pre-processing technique to deal with large flight data with high frequencies, as described in [19], [20]. Therefore, in this work, time-series resampling technique is also adopted to transform the flight data from every 60 seconds to one data point by down-sampling, while more data points are interpolated into the original weather data by up-sampling. Resampling also ensures that the multivariate time-series data runs continuously at a standardised frequency over the data period, compensating for erroneous data due to missing or duplicate values. Subsequently, spatial alignment is performed using the k-Nearest Neighbours (k-NN) algorithm. Given that there are only three weather stations spaced roughly equally apart, the 1-nearest neighbour approach ($k=1$) is employed. In this method, the algorithm identifies the closest point from the flight data to weather data, at each date-time index, using the Euclidean distance. By setting $k=1$, only the nearest data point is selected for alignment, which simplifies the process and ensures efficiency. This approach provides accurate spatial alignment with minimal computational overhead, as only the single closest neighbour is considered.

Flight data is matched with the weather data from the nearest station at each date-time index to derive the combined environmental and operational data. The initial combined EOC data will have missing values in the flight data during non-operational phases. Operational variables such as Indicated Airspeed (IAS), accelerations, pitch and roll movements, can be assumed as zero. Geographical information would be based on the last available data as we assume the aircraft to be stationary during its non-operational phase. Environmental variables, such as Outside Air Temperature (OAT) would have to be estimated from other known data. In this instance, OAT is estimated by employing the relationship between the temperature from the weather data and the OAT from the aircraft's OAT sensor when $IAS = 0$ (i.e. aircraft is powered up on ground and stationary). Detailed results are shown and discussed in the next section.

Fig. 1: Data fusion framework



Case Study and Results

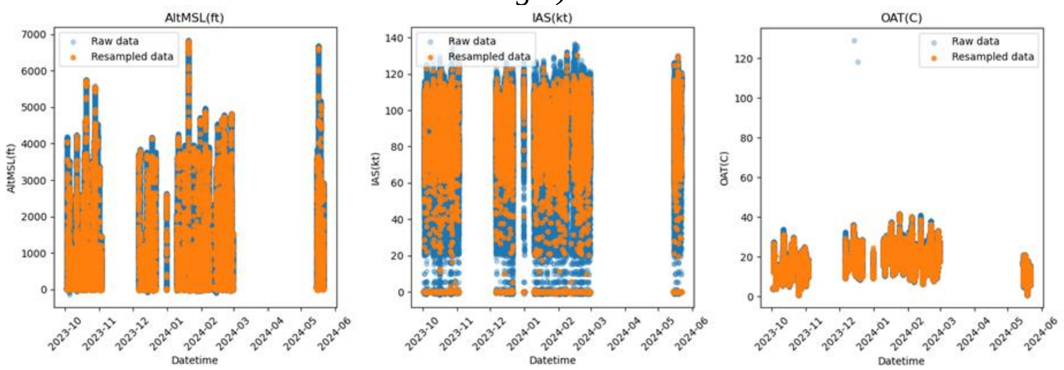
The case study for applying the proposed framework uses approximately 6 months of data (December 2023 to May 2024) collected from the respective weather stations and a selected operational Cessna 172S typically used for flight training between Bendigo, Point Cook, and Moorabbin. The raw flight data contains 614,010 rows of parameters across this period. A few down-sampling methods were attempted, including 1) retaining the first value of each minute's data, 2) obtaining the median of each minute of data, and 3) obtaining the median of data smoothed over a rolling window of 30 data points (i.e. 30 seconds). Each method was benchmarked against the raw data distribution using the Kolmogorov-Smirnov test (K-S test) and the Anderson-Darling test (AD test). The K-S test is a statistical method to compare the distribution of two datasets for similarity. Similarly, the AD test also checks for similarity in distributions between two datasets, but puts more weight on the extreme tails of the distributions. Fig. 2 shows the p-value of the statistical tests between different down-sampling methods. Assuming a standard 95% confidence level, the null hypothesis that the samples come from the same underlying distribution can be rejected where *p-value* ≤ 0.05 (highlighted in green in Fig. 2). Based on the results of the statistical tests, method 1 is adopted for latitude, longitude, IAS, accelerations, pitch, and roll; while method 3 is adopted for altitude above mean sea level (AltMSL), OAT, and pressure. Fig. 3 shows selected resampled and raw flight parameters scatter plotted against time for comparisons.

Fig. 2: Flight Data Resampling Results

Down-sampling methods	Test	Latitude	Longitude	AltMSL(ft)	IAS(kt)	OAT(C)	Pressure (inHg)	NormAc(G)	LatAc(G)	Pitch	Roll
1) Retaining first value	A-D	0.25	0.25	0.0376	0.0762	0.25	0.25	0.0552	0.25	0.25	0.088
	K-S	0.137	0.493	0.0269	0.198	0.962	1	0.339	0.835	0.534	0.141
2) Median	A-D	0.25	0.25	0.125	0.001	0.25	0.25	0.001	0.001	0.001	0.001
	K-S	0.171	0.0576	0.01	0	0.985	1	0	0	0	0
3) Median of smoothed data	A-D	0.25	0.25	0.0531	0.121	0.25	0.25	0.001	0.001	0.112	0.001
	K-S	0.167	0.552	0.0388	0.208	0.984	1	0	0	0.383	0.301
Selected method for each parameter		1	1	3	1	3	3	1	1	1	1
Combined	A-D	0.25	0.25	0.071	0.1	0.25	0.25	0.055	0.25	0.25	0.088
	K-S	0.16	0.53	0.05	0.23	0.94	1	0.34	0.83	0.53	0.14

Weather data is interpolated by assuming linearity between the approximate 30-minute intervals. The authors note that this is a limitation of the methodology adopted because linear interpolation fails to capture the variances and sporadic nature of the weather. However, given that the objective of the model is to analyse the impact of EOC on long-term structural degradation such as corrosion, it is reasonable to assume that the minute variations in environmental conditions would have a less significant effect than the interactions during operational phases. Therefore, while linear interpolation may miss short-term fluctuations, it remains sufficient for modelling the long-term environmental exposure relevant to structural health monitoring (SHM).

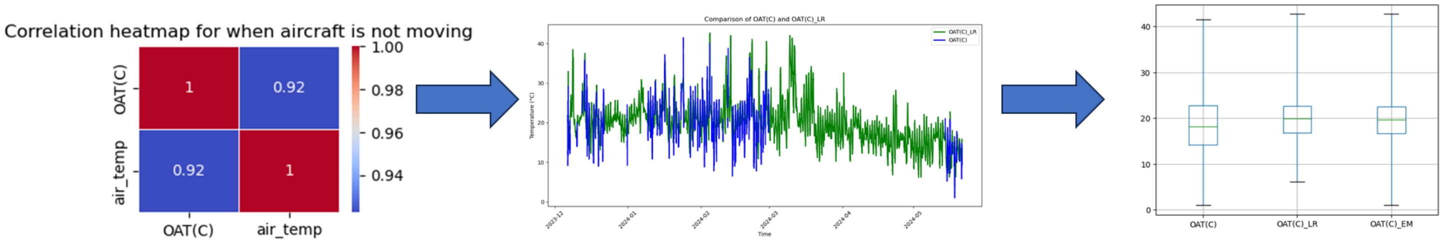
Fig. 3: Resampled and raw (a) AltMSL, (b) IAS, and (c) OAT plotted against time (from left to right)



Subsequently, as discussed earlier, the OAT in the combined EOC dataset contains missing values during non-operational phases. A strong linear correlation can be observed between the OAT and the air temperature obtained from the weather stations when the aircraft is on the ground (i.e. *IAS = 0*) based on the *Pearson correlation coefficient = 0.92*. Hence, to shape a complete picture of the historical representation of the temperature around the aircraft, a Linear Regression (LR) model coupled with an Expectation-Maximization (E-M) algorithm is used to estimate the missing values from known parameters (Fig. 4). It should be noted that this method adopted to impute and estimate the missing values is incumbent on the strong linear correlation between the available data, which may not always hold especially for variables with limited operational data, such as the relative humidity and wind speed. These will need additional computations that are not discussed in the context of this paper.

In the Expectation step, the non-missing values during the operational phase are used to initialise the algorithm, thereby retaining actual OAT values, before filling the missing values with the trained LR model. Subsequently, all values are refitted into the LR model to refine the values where data was previously missing. As depicted in the boxplots in Fig. 4, it can be observed that the E-M algorithm brings the overall median, minimum and maximum values closer to that of the original parameter.

Fig. 4: Expectation-Maximisation process explained



Expectation step:
Fill missing values with Linear Regression (LR) model

Maximisation step:
Refit LR model with filled values for refinement

Since there is no ground truth to assess the accuracy of the LR model and E-M algorithm, the models were applied to estimate the OAT during the operational phase (i.e. training dataset) and compared with the actual OAT values for validation. Results showed that the E-M algorithm improves the estimation of the missing parameters.

Table 1: Error metrics table of the LR model only and LR model with E-M algorithm tested on the training dataset

Error metrics	LR model only	LR model with E-M algorithm
Mean Absolute Error (MAE)	4.74	3.49
Root Mean Squared Error (RMSE)	6.01	4.27

Potential Application Areas

The proposed spatiotemporal data fusion technique completes the understanding of both operational and non-operational phases an aircraft experiences throughout its service life, thereby improving aircraft lifecycle management processes. Moreover, the combination of spatial information and time-series data can feed an aircraft digital twin to represent the physical aircraft in its dynamic operating environment virtually. Such an approach could potentially be applied not only to PHM and SHM as discussed in the earlier sections, but also support other areas of aviation research including, but not limited to, flight path optimisation for extreme weather avoidance, fuel efficiency and emissions reduction.

Conclusion

This paper introduces a spatiotemporal data fusion method to address the gap in capturing the varying EOC throughout the aircraft's lifecycle. By integrating flight data from the Garmin G1000 system with weather data, the presented approach constructs a comprehensive EOC profile for a Cessna 172S aircraft without additional onboard monitoring sensors. Temporal resampling and spatial k-nearest-neighbour methods ensure data alignment, while the complete representation of EOC enables more holistic health and usage monitoring. This paves the way for future works to incorporate the combined EOC data with data-driven predictive models to develop an aircraft digital twin for structural prognostics and health management.

References

[1] I. G. Hebden, A. M. Crowley, and W. Black, 'Overview of the F-35 structural prognostics and health management system', presented at the Proceedings of 9th European workshop on SHM, July, 2018, pp. 10–13.

[2] T. Dong and N. Kim, 'Cost-Effectiveness of Structural Health Monitoring in Fuselage Maintenance of the Civil Aviation Industry †', *Aerospace*, vol. 5, no. 3, p. 87, Aug. 2018, doi: 10.3390/aerospace5030087.

[3] G. Li, H. Lee, A. Rai, and A. Chattopadhyay, 'Analysis of operational and mechanical anomalies in scheduled commercial flights using a logarithmic multivariate Gaussian model', *Transportation Research Part C: Emerging Technologies*, vol. 110, pp. 20–39, Jan. 2020, doi: 10.1016/j.trc.2019.11.011.

[4] Y. Zhang, S. Cao, B. Wang, and Z. Yin, 'A Flight Parameter-Based Aircraft Structural Load Monitoring Method Using a Genetic Algorithm Enhanced Extreme Learning Machine', *Applied Sciences (Switzerland)*, vol. 13, no. 6, Mar. 2023, doi: 10.3390/app13064018.

[5] E. Q. Wu *et al.*, 'Flight Situation Recognition Under Different Weather Conditions', *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 3, pp. 1753–1767, Jun. 2021, doi: 10.1109/TAES.2020.3048777.

[6] C. Keryk, R. Sabatini, K. Kourousis, A. Gardi, and J. M. Silva, 'An Innovative Structural Fatigue Monitoring Solution for General Aviation Aircraft', *J.Aerosp. Technol. Manag.*, vol. 10, Feb. 2018, doi: 10.5028/jatm.v10.779.

[7] T. G. Puranik and D. N. Mavris, 'Identification of Instantaneous Anomalies in General Aviation Operations Using Energy Metrics', *Journal of Aerospace Information Systems*, vol. 17, no. 1, pp. 51–65, Jan. 2020, doi: 10.2514/1.I010772.

[8] D. Zhou, X. Zhuang, H. Zuo, H. Wang, and H. Yan, 'Deep Learning-Based Approach for Civil Aircraft Hazard Identification and Prediction', *IEEE Access*, vol. 8, pp. 103665–103683, 2020, doi: 10.1109/ACCESS.2020.2997371.

[9] S. Pfingstl, D. Steinweg, M. Zimmermann, and M. Hornung, 'On the Potential of Extending Aircraft Service Time Using Load Monitoring', *Journal of Aircraft*, vol. 59, no. 2, pp. 377–385, 2022, doi: 10.2514/1.C036569.

[10] S. Ignatovich, S. Yutskevych, and I. Makarov, 'Features of aircraft structures corrosion damage during storage under COVID-19 restrictions', *Procedia Structural Integrity*, vol. 36, pp. 66–70, 2022, doi: 10.1016/j.prostr.2022.01.004.

[11] E. J. Tuegel, A. R. Ingraffea, T. G. Eason, and S. M. Spottswood, 'Reengineering Aircraft Structural Life Prediction Using a Digital Twin', *International Journal of Aerospace Engineering*, vol. 2011, pp. 1–14, 2011, doi: 10.1155/2011/154798.

[12] M. Liao, G. Renaud, and Y. Bombardier, 'Airframe digital twin technology adaptability assessment and technology demonstration', *Engineering Fracture Mechanics*, vol. 225, p. 106793, Feb. 2020, doi: 10.1016/j.engfracmech.2019.106793.

[13] G. M. Bisanti, L. Mainetti, T. Montanaro, L. Patrono, and I. Sergi, 'Digital twins for aircraft maintenance and operation: A systematic literature review and an IoT-enabled modular architecture', *Internet of Things*, vol. 24, p. 100991, Dec. 2023, doi: 10.1016/j.iot.2023.100991.

[14] A. Sadeghi, P. Bellavista, W. Song, and M. Yazdani-Asrami, 'Digital Twins for Condition and Fleet Monitoring of Aircraft: Towards More-Intelligent Electrified Aviation Systems', *IEEE Access*, pp. 1–1, 2024, doi: 10.1109/ACCESS.2024.3371902.

[15] B. Wang, J. Sun, D. Chen, and S. Ning, 'A cumulative exposure model for remaining useful life prediction of pitot tube based on dynamic operating data', *Quality and Reliability Engineering International*, vol. 39, no. 4, pp. 1518–1531, 2023, doi: 10.1002/qre.3299.

[16] A. Trueman, A. Butler, J. Hopf, and J. Jacob, 'Field testing of a corrosion prediction health management system on military aircraft', in *Corrosion and Prevention 2014*, Darwin, Australia: Australasian Corrosion Association, Sep. 2014.

[17] N. Fala, G. Georgalis, and N. Arzamani, 'Study on Machine Learning Methods for General Aviation Flight Phase Identification', *Journal of Aerospace Information Systems*, vol. 20, no. 10, pp. 636–647, Oct. 2023, doi: 10.2514/1.I011246.

[18] Bureau of Meteorology (BoM), 'About latest weather observations', About latest weather observations. Accessed: Sep. 21, 2024. [Online]. Available: http://www.bom.gov.au/catalogue/observations/about-weather-observations.shtml

[19] M. B. Slimene and M.-S. Ouali, 'Anomaly Detection Method of Aircraft System using Multivariate Time Series Clustering and Classification Techniques', in *IFAC-PapersOnLine*, Nantes, France: Elsevier, 2022, pp. 1582–1587. doi: 10.1016/j.ifacol.2022.09.616.

[20] E. J. Garcia, M. L. Mulvihil, M. S. Kharab, C. L. Stephens, and N. J. Napoli, 'Capturing Complex Multivariate Time Series Interactions to Detect High-Risk Adverse Events During Flight', in *AIAA 2023-3548*, San Diego, CA and Online: AIAA, 8 June 20233. doi: https://doi.org/10.2514/6.2023-3548.