


# Data augmentation for predictive maintenance: Synthesising aircraft landing gear datasets

Izaak Stanton<sup>1</sup>  | Kamran Munir<sup>1</sup> | Ahsan Ikram<sup>1</sup> | Murad El-Bakry<sup>2</sup>

<sup>1</sup>Computer Science Research Centre (CSRC), School of Computing and Creative Technologies (SCC), College of Arts, Technology and Environment (CATE), University of the West of England (UWE), Bristol, UK

<sup>2</sup>Airbus Operations Ltd. Pegasus House, Aerospace Avenue, Filton, UK

## Correspondence

Izaak Stanton, Computer Science Research Centre (CSRC), School of Computing and Creative Technologies (SCC), College of Arts, Technology and Environment (CATE), University of the West of England (UWE), Bristol, BS16 1QY, UK.

Email: [izaak3.stanton@uwe.ac.uk](mailto:izaak3.stanton@uwe.ac.uk)

## Abstract

In the aviation industry, predictive maintenance is vital to minimise Unscheduled faults and maintain the operational availability of aircraft. However, the amount of open data available for research is limited due to the proprietary nature of aircraft data. In this work, six time-series datasets are synthesised using the DoppelGANger model trained on real Airbus datasets from landing gear systems. The synthesised datasets contain no proprietary information, but maintain the shape and patterns present in the original, making them suitable for testing novel PdM models. They can be used by researchers outside of the industry to explore a more diverse selection of aircraft systems, and the proposed methodology can be replicated by industry data scientists to synthesise and release more data to the public. The results of this study demonstrate the feasibility and effectiveness of using the DoppelGANger model from the Gretel.ai library to generate new time series data that can be used to train predictive maintenance models for industry problems. These synthetic datasets were subject to fidelity testing using six metrics. The six datasets are available on the UWE Library service.

## KEYWORDS

aircraft maintenance, generative adversarial network, machine learning, predictive maintenance, synthetic data

## 1 | INTRODUCTION

Over the past two decades, most aircraft accidents occurred during the take-off and landing flight phases.<sup>1</sup> These phases at the start and end of flights rely on not only the skill of the pilot but on the health and quality of the aircraft's landing gear, the failure of which made up to 60% of all recorded aircraft problems in 2012.<sup>2</sup> Airlines implement maintenance, repair and operation (MRO) to mitigate these accidents, and ensure the quality, safety and health of their aircraft, with \$69 billion spent globally on aircraft maintenance in 2018 alone.<sup>3</sup> There has been a gradual shift from reactive to proactive maintenance for aircraft, with predictive maintenance (PdM) gaining prominence as a vital practice in the aviation industry. PdM focuses on analysing aircraft data to predict when and where systems will fail, assessing the health of aircraft systems, and enabling proactive measures to maintain them. The goal of PdM is to reduce aircraft downtime,<sup>4</sup> in turn saving the airlines money with precision optimal maintenance and reduces the number of unscheduled faults and failures. It however requires an experienced staff to build PdM models well suited for specific aircraft systems and relies

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Engineering Reports* published by John Wiley & Sons, Ltd.

heavily on a large quantity of high-quality data, which just is not always available within the industry, and even less so to the public.

This lack of available data limits the research possible to academics outside of airlines and manufacturers, who face many challenges in acquiring the data to build novel models. Firstly, to ensure the safety of passengers and crew and the aircraft's operational availability, aircraft systems regularly undergo MRO long before they can fail, resulting in a lack of failure data to train models on. Secondly, in the public sector, this data is entirely unavailable due to the proprietary nature of airline data. This limits researchers to using simulated datasets like Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) datasets.<sup>5</sup> While these datasets are popular, researchers like Xiong et al. agree that datasets such as these are inaccurate, and that papers that use alternative data do not disclose it.<sup>6</sup> Finally, these simulated datasets cover a minimal scope of aircraft systems. In a previously conducted state-of-the-art review, it was found that there was a large bias towards turbofan engines for testing novel PdM models.<sup>7</sup> Equally vital components such as landing gear or fuselage are overlooked due to the lack of readily available data outside of the industry.

One solution to mitigate these problems is to synthesize new datasets, using models trained on data from existing datasets provided by aircraft manufacturers. This would produce data that preserves the shape and patterns exhibited by the original while anonymising it to be used more easily by researchers unaffiliated with manufacturers. In the last five years, there have been many different models proposed for synthesising time series data, which is already being used for automotive<sup>8</sup> and electric vehicles.<sup>9</sup> At the time of writing this paper, no example of synthetic aircraft data that was released publicly using models trained on real datasets could be found. This paper focuses on utilising real industry data of aircraft landing gear to train synthetic models using the DoppleGANger model in Gretel.ai to generate six new synthetic datasets for aircraft landing gear subsystems.

## 1.1 | Structure

The structure of this paper is as follows. Section 2 covers related works including what models are available for synthesising time series data and where they are being used in similar industries. Section 3 outlines the research methodology for this research, including what metrics the quality of the synthetic data was measured against. Section 4 reviews datasets, describing their primary features, shape and required preprocessing. Section 5 displays the results of the trained models against the selected validation metrics, and contains suggestions for how the synthetic data can be utilised. Finally, Section 6 contains the conclusions that can be drawn and ideas for future work.

## 1.2 | Contributions

The primary contribution of this paper is the six synthetic landing gear datasets built from models trained from existing data. With sufficient validation, these can be made available to the public for open use to provide data for a mix of regression and classification problems. The non-proprietary data produced for this paper was published online by the \*University of the West of England's Research Data Repository. These datasets will benefit researchers without access to industry datasets to explore a more diverse range of aircraft systems for new PdM models. By demonstrating the feasibility of Gretel.ai, the methodology of this paper can be used by researchers within the industry to produce a greater number of datasets in the future. These datasets could be used in conjunction with the original, to increase the size of the training dataset to improve performance, which will be explored further in the future.

## 2 | RELATED WORK

Synthetic data refers to artificially generated datasets that mimic the characteristics and statistical properties of real-world data. Synthetic data can augment existing data to increase the size of training datasets. This can improve the performance of machine learning models, or counter data scarcity where less data exists for certain classes. Synthetic data also preserve the privacy of the original and can be generated without the need for complex simulators or tools. One challenge is that Synthetic data only resembles the data it was trained on however, so outliers and anomalous results that did not occur in the original, or even just not frequent enough to accurately be modelled will not be synthesised correctly.

Synthetic data can be generated using traditional models such as linear regression for Reference 10, and decision trees for Reference 11. More recently machine learning algorithms have been employed to synthesise new data, particularly Recurrent Neural Networks (RNN) and Generative Adversarial Networks (GAN). As the need for deep learning grows, so does the need for greater data, with different fields where privacy is of the utmost importance benefiting from data synthesised, such as healthcare data using models such as SynSyn<sup>12</sup> or, financial data.<sup>13</sup>

Synthetic data is already being used to generate data with greater privacy using recent advances in image generation. To address privacy and ethical concerns of using face image datasets of real people,<sup>14</sup> used StyleGAN2-ADA to generate their own dataset of synthetic face images named SFace. This was tested with multi-class classification models to attain 99.13% validation accuracy and released the code and face image dataset publicly on GitHub.<sup>15</sup> Developed a data collector and labeller to generate synthetic crowd images. They then improved the performance of crowd counting by pretraining the crowd counter using the synthetic data, and then fine-tuning it using the real data, achieving higher than state-of-the-art performance. They also released the source code and datasets on GitHub.<sup>16</sup> generated datasets of 52,500 images for UAV detection in images, and also fine-tune the models using real-world data near the end of training. They improved the performance of aerial object detection under varying illumination conditions and determined the shape of the UAV is more important for image detection than the texture.

Since their inception in 2014,<sup>17</sup> Generative Adversarial Networks (GANs) have been used to generate new instances of data, including images, music and synthetic data. The purpose of a GAN is to generate data similar to that in the training datasets using unsupervised learning. GANs traditionally consist of a generator to generate data, and a discriminator to determine how realistic it is against a training dataset. This way GANs can generate realistic synthetic data and excel at generating data where high realism is required, such as computer vision and natural language processing. GANs have shown promising results in synthesising images, audio, and text data, and their use has become widespread since their proposal. Several novel variations of GANs have appeared in recent years for synthesising data, such as COT-Gan for generating sequential data,<sup>18</sup> or image generation with DCGAN<sup>19</sup> and StackGAN.<sup>20</sup> Time series data have temporal dependencies that traditional GANs do not consider. Several models designed specifically to handle realistic time series generation based on GANs have been developed. A systematic literature review was performed in Reference 21. The paper highlights the challenges with Time Series GANs, maintaining training stability, difficulties in evaluating synthesised data, the risk of a privacy breach and the lack of benchmark time series datasets. They also list a comprehensive list of applications, architectures and evaluation metrics. RCGAN<sup>22</sup> produces real-value multi-dimensional time series data for medical data, using RNNs in the generator and discriminator condition on auxiliary information. A novel evaluation measure was tested against the MNIST hand-written digit dataset,<sup>23</sup> treating the  $28 \times 28$  pixel images as a 28-parameter time series across 28 timesteps, as it could be visually checked for similarity<sup>24</sup> developed TimeGAN a network consisting of embedding, recovery, sequence generator and sequence discriminator, simultaneously learning to encode features, generate represent and iterate across time from the jointly trained adversarial components. This model aims to preserve the temporal dynamics of the original data, and qualitatively and quantitatively outperformed benchmarks. The codebase released on GitHub is currently outdated, but it is found in other open-source libraries such as ydata-synthetic<sup>†</sup>. DoppelGANger<sup>25</sup> has a focus on metadata, achieving a 43% better fidelity than the baseline. The model's only drawback is not resolving the privacy problem, however, it is used in open-source libraries such as Gretel-synthetics on GitHub<sup>‡</sup> with a responsive community on Discord. PART-GAN is another GAN, that prioritises preserving the privacy of the original data.<sup>26</sup> Compared to other models, it can enable the generation of an unlimited amount of data, provides robust privacy and addresses the trade-off between utility and privacy using optimisation strategies.

### 3 | METHOD

The methodology used for this paper is as follows:

1. Select Industry datasets fitting the research scope and model requirements.
2. Clean the data, removing anomalies not representative of the dataset.
3. Shape the data to  $(X, Y, Z)$  dimensions:  $X$  datasets,  $Y$  timesteps,  $Z$  features.
4. Train the chosen model for time series synthesis.
5. Evaluate the performance of each model using fidelity testing metrics.
6. Optimise the hyperparameters of the model using Bayesian optimisation provided by Optuna.
7. Generate final datasets using optimised models.
8. Gather final results using fidelity metrics and visual inspection.

### 3.1 | Research scope

The scope of the datasets used for training the GAN in this paper is outlined in the Venn diagram shown in Figure 1. As the main goal of this paper is to produce datasets of similar quality to proprietary datasets from aircraft manufacturers, only the industry datasets provided will be used to train the synthetic data model. These were taken directly from data recorded by in-service aircraft manufactured by Airbus. This ensured the synthesised data reflects real-world data already being analysed for PdM tasks. Most data recorded in aircraft is time series, with recordings at regular intervals during flight, or at the start and end of specific flight phases. Therefore, the data type selected for use in this study was time series data, with each time step representing either a flight cycle or a regular time step. Finally, there are a myriad of complex aircraft systems that produce data and are vital to maintaining the operational availability of the aircraft, this paper will focus on one of the most important, aircraft landing gear. All the training data sets were sourced from aircraft components that make up or relate to aircraft landing gear. With this research scope established, the next step is to collect data that adheres to it.

### 3.2 | Data gathering

Airbus provided the data used in this research, collected from Aircraft Condition Monitoring Systems (ACMS), a PdM enabling tool that consists of a high-capacity flight data unit and the associated sensors.<sup>27</sup> These datasets vary in monitoring period, ranging from the past 2–6 years. The length of each dataset varies based on the starting date of the recordings. Therefore to maximise the number of training datasets available, the longer datasets were split down to size, the details of which are described for each dataset. All the datasets were selected from components related to aircraft landing gear belonging to Airbus aircraft. The location of where these aircraft were operated, their identifiers and the names of the airlines using them have been removed, but they vary greatly across the datasets unless specified otherwise. The maintenance protocols for each airline will differ in small ways, and aircraft in different geographical locations will encounter different environmental factors affecting their operability. Each dataset varies in data gathering regime, set of sensors, frequency of recording rate and operator environment. Of all the identified datasets, none had pre-established classes for anomalous or failure data, limiting the analysis to unsupervised learning models. More details for how the real industry datasets were selected are given at the start of Section 4.

### 3.3 | Preprocessing

For the most part, the data was left aside from some rare instances of removing null values. Fundamentally the synthetic data should maintain all the characteristics of the original. For example, noise in several noisier datasets could not be removed, as the time series may fall below a threshold value from simply a noise spike. Even in this case, it would be a

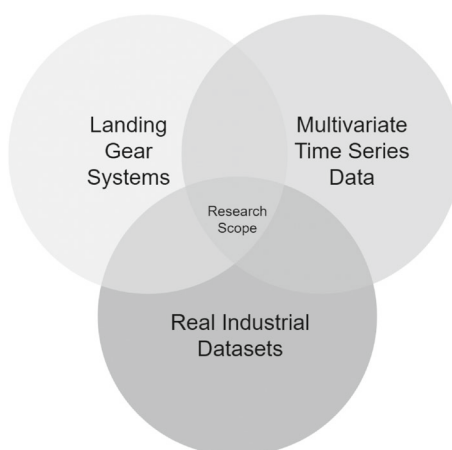


FIGURE 1 Venn diagram highlighting the research scope of this paper.

protocol to consider that component approaching failure even if it doesn't track with the average decline, and maintenance would be conducted prematurely. It is up to the users of the synthetic datasets to clean the synthetic data themselves as they see appropriate.

The DoppelGANger model requires the data to be shaped as a three-dimensional Numpy array. The original datasets were reshaped into shape  $(X, Y, Z)$  where the  $X$  axis is for the unique sets of data, the  $Y$  axis is for the time step, and the  $Z$  axis is for the feature. For example,  $(50, 100, 8)$  would be 50 unique sets of data, each of length 100 time steps, with eight different features. All the original datasets are time series with multiple features, so splitting to  $Z$  features is simple. The size of  $X$  and  $Y$  for each dataset was balanced to maximise the number of datasets for the training data to improve the overall performance of the model while maintaining long enough sets of data to be usable for PdM problems. This was done on a case-by-case basis and more details discussing this are discussed in Sections 4.1–4.6.

### 3.4 | Synthetic method selection

A suitable model to synthesise the data was required, one that maintains the same relationship and properties of the original without revealing the original dataset. The model must be able to produce enough variety to create enough datasets for implementing a supervised learning approach for both regression and anomaly detection problems. The choice of model for synthesising data depends on the nature of the data and the objectives of the research. For this research, the model must be capable of synthesising time series data of up to 200 elements long, as well as being readily available with ongoing support to new researchers.

Gretel.ai's *gretel-synthetics* library was used as the library of choice for implementing DoppelGANger, offering several advantages for this project. Firstly, it is an open-source library, providing flexibility, support from the community and developers, and availability for reproducing the results. Additionally, an element of Gretel.ai's novelty is incorporating metadata such as the location, and then training a model. While not in use for the data in this research, it would prove valuable for future analyses. Knowing the airport's source and target allows for location, and even the environment and weather by referencing available data from Meteorological Aerodrome Reports (METAR) recorded at respective airports.<sup>28</sup> This integration enables a comprehensive understanding of the synthesised time series data. Therefore, Gretel.ai was used to generate synthetic data for these datasets.

### 3.5 | Evaluation and fidelity testing

The world is entering an era where synthetic and hybrid datasets are a necessity to augment existing datasets required for deep learning tools. The validation of the synthetic data is just as vital, to ensure that it maintains the same recognisable shape and patterns as the original, without containing any exact duplicate of the data. For this research, there are three primary requirements that the synthetic data must meet to be successfully validated.

1. The synthesised data is indiscernible to the average viewer from the original data. Therefore the synthetic data had to maintain the same overall shape, range and patterns.
2. No part of the original data is directly present in the synthesised data to maintain the airline's privacy.
3. In recent years, several qualitative and quantitative evaluation techniques have been employed and developed to successfully confirm the fidelity of the dataset. The following six metrics were used to validate the synthetic data.

The fidelity metrics used for this evaluation are as follows.

#### 3.5.1 | Autocorrelation

Autocorrelation measures the similarity between a time series and a shifted version of itself at different time lags, indicating the presence and strength of any repeating patterns or trends in the data. It has been used to assess whether original and synthetic data maintain the same shape and patterns in several recent publications. The two sets of data with the least mean squared error (MSE) was selected and displayed alongside autocorrelation plots. These plots reveal the

presence of any temporal dependencies or trends shared by either dataset. Comparing the autocorrelation patterns allows us to determine whether the synthetic data captures the temporal dynamics exhibited by the original dataset.

### 3.5.2 | Pearson correlation heatmap

The Pearson correlation coefficient can be used to quantify the linear relationships between variables in a dataset, and has been widely employed in journals to analyse feature correlations. This method compares the features of the original and synthetic datasets, to ensure they maintain the same correlation. A heatmap of correlations is generated using this technique, with each row and column representing a different feature. The average correlation across each instance in the original and synthetic datasets is then calculated, and the difference between them is used to populate the heatmap. The aim is to minimize the difference and prove a similar correlation between the features, with a universal value near zero indicating similar correlations in both the original and synthetic datasets. This method does require the assumption that the features are not independent of each other.

### 3.5.3 | Principle component analysis

Principle component analysis (PCA) is employed as a dimensionality reduction technique, which operates particularly effectively with the linear data that the Airbus datasets exhibit. By analysing the principal components of both the original and synthetic data, the similarity in their underlying structure can be assessed. If the shape and spread of the synthetic PCA match the original, it indicates that the model has successfully captured the essential underlying structure and variability present in the original dataset. An alternative method, the more recently introduced T-SNE which visualizes and compare the data's high-dimensional relationships, could have also been used. This is a nonlinear dimensionality reduction method however, and as most of the data in these datasets are reasonably linear, the results are less likely to provide any meaningful data.

### 3.5.4 | Kullback Leibler

The Kullback-Leibler (KL) divergence is a mathematical measure that quantifies the distance between two probability distributions, where a value of 0 indicates that the distributions are identical. The equation for calculating the divergence between two probability distributions,  $P$  and  $Q$ , is shown in the following equation.

$$D_{KL}(P|| Q) = \sum_i P(i) \cdot \log\left(\frac{P(i)}{Q(i)}\right) \quad (1)$$

By converting the training and original datasets into discrete probability distributions, the KL divergence can gauge the similarity between their underlying data distributions. The mean and standard deviation of the KL divergence across a random selection of training and synthetic instances were calculated, indicating the overlap between the dataset's distributional patterns. This approach enables a more nuanced assessment of the synthesised data's fidelity in capturing the essential characteristics of the original dataset.

### 3.5.5 | Train on synthetic, test on real

The Train on Synthetic Test on Real (TSTR) is a quantitative analysis approach to train a prediction model using synthetic data and test it against the original data. The performance of this model can then be compared to one trained on real, and tested on real (TRTR). For this evaluation, an LSTM model is trained using a TensorFlow sequential model to predict the next step in the time series, with a random selection of datasets used for training, 25% of which was reserved for testing. Additionally, a hybrid training setup can be explored by training on both synthetic and real data and testing solely on the real data (TTSTR) however this will be explored in future work against Airbus's previous methods for predictive and classification models. This method aims to evaluate if a combined training dataset improves overall model performance.

### 3.5.6 | Time series classification

The synthetic data must be indistinguishable from the original data, and one last way to assess this is to use a classification model and compare the performance of it between the original and synthetic data. If both datasets achieve high performance when input into the model, they must both possess the same temporal properties making them difficult to distinguish. A 2-layer LSTM model is trained using a TensorFlow sequential model to classify whether the input data originated from the original dataset. The model is trained until it reaches a suitable performance against a test dataset of the original data, measured on its precision and accuracy. The synthetic data is then input into the model, and the performances can be compared.

### 3.6 | Hyperparameter optimisation

Once the training data was reshaped, and the performance metrics were tested, the hyperparameters of the models were optimised using Optuna.<sup>29</sup> Optuna is an open-source optimisation framework for automated hyperparameter searching. The Bayesian optimization it provided operated more efficiently to build the ideal model than a standard grid or random search. Trials were conducted using a set of possible values for each hyperparameter selected through experimentation, with each trial attempting a new combination of the hyperparameters. Some models took longer to train than others due to the size of the training data, so each model was run through as many trials as possible in 48-h time slots for each model, with no real improvement in performance witnessed in training runs longer than this. At the end of each training run, the model with the lowest Kullback Leibler score was highlighted, and the remaining performance metrics were applied to confirm it performed well across every metric. The optimisation tool and the hyperparameters of the model to tune were both recommended by members of the Gretel.ai team.

## 4 | REAL INDUSTRY DATASET OVERVIEW

In the pursuit of creating acceptable synthesized data for PdM problems, industrial datasets already being investigated for PdM models were chosen. Provided the synthetic data maintains the same temporal behaviour, the output can be used for the same PdM testing. Following an evaluation of Airbus-provided datasets, six were chosen for their diversity in landing gear systems and ML problem types. The selection process for each dataset adhered to the following guidelines:

1. The training data is assumed to be clean, and any visually anomalous data will be removed to produce a dataset of “typical” training data.
2. The shape, that is the visual and statistical characteristics describing the overall pattern, trend, and seasonality, of each set of training data is determined by the key parameters and the time length specified in Table 1.
3. There are no exact dates or times for each data point representing either a flight cycle or a time increment in live monitoring, specified in Table 1.

**TABLE 1** An overview of the six industrial datasets used for training.

Dataset identifier	System	Size	Parameters	Problem type
Training Dataset 1	Bogie Pitch Trimmer	[175, 200, 8]	Oil Pressure	Regression
Training Dataset 2	Tyre Pressure Indication System	[40, 200, 8]	Tyre Pressure	Regression
Training Dataset 2	Tyre Pressure Indication System	[9283, 180, 2]	Tyre Pressure, Brake Temperature	Classification
Training Dataset 3	Landing Gear Brakes	[194, 100, 2]	Brake Temperature, Wheel Speed	Both
Training Dataset 5	Landing Gear Brakes	[109, 100, 24]	Brake Temperature, Tyre Pressure	Classification
Training Dataset 6	Landing Gear Brakes	[37, 50, 24]	Brake Temperature, Deceleration	Classification

Note: Size is formatted as [X, Y, Z] where X is the number of sets of data, Y is the number of time steps and Z is the number of parameters.

Sections 4.1–4.6 give a detailed description of each training dataset. An overview of the details of each dataset is also summarised in Table 1. Note that Figure 2E only display plots for 6 of their 24 parameters for ease of reading.

#### 4.1 | Training dataset 1: Bogie pitch trimmer

All aircraft incorporate hydraulically powered components, mechanisms that employ pressurized fluid to drive machinery and facilitate the movement of mechanical elements such as brakes, flaps and landing gear. Over the course of many flight cycles, there is a gradual loss of hydraulic fluid, necessitating maintenance to ensure that the fluid level remains above a critical threshold, preventing potential system failures. The first dataset is a multivariate time series dataset for a Bogie Pitch Trimmer, a hydraulic system responsible for adjusting the angle of the aircraft's landing gear assemblies. The dataset measures the decline in the system's hydraulic fluid caused by leakage between flight cycles, across a selection of aircraft of the same design. The primary parameters of this dataset are the oil pressure in the left and right landing gear, which decreases in line with the fluid level. The dataset consists of 175 sets of data, each 200 flight cycles long, across 8 different parameters. The original datasets all contained 200 cycles of data or more but any exceeding 400 were cut down to increase the size of the overall dataset to improve the performance of the model during training. This data is used to predict the RUL of the system, with regression models predicting the decline of the pressure until it hits a known critical threshold. Only the last 100 flight cycles are required to predict this, so the decrease in size does not affect its use in PdM testing. Each dataset represents a slight general decline, with a lot of noise caused by static affecting the readings and environmental factors (Ambient temperature, altitude of airport etc.). Some datasets will contain sharp spikes in pressure where the fluid has been refilled. Figure 2A shows an example of this data.

#### 4.2 | Training dataset 2: Tyre pressure indicator

Ensuring proper tyre pressures through regular maintenance is crucial to alleviate overheating and wear problems. These issues, if left unattended, can result in tyre deflation or even explosive break-up during critical flight phases like landing

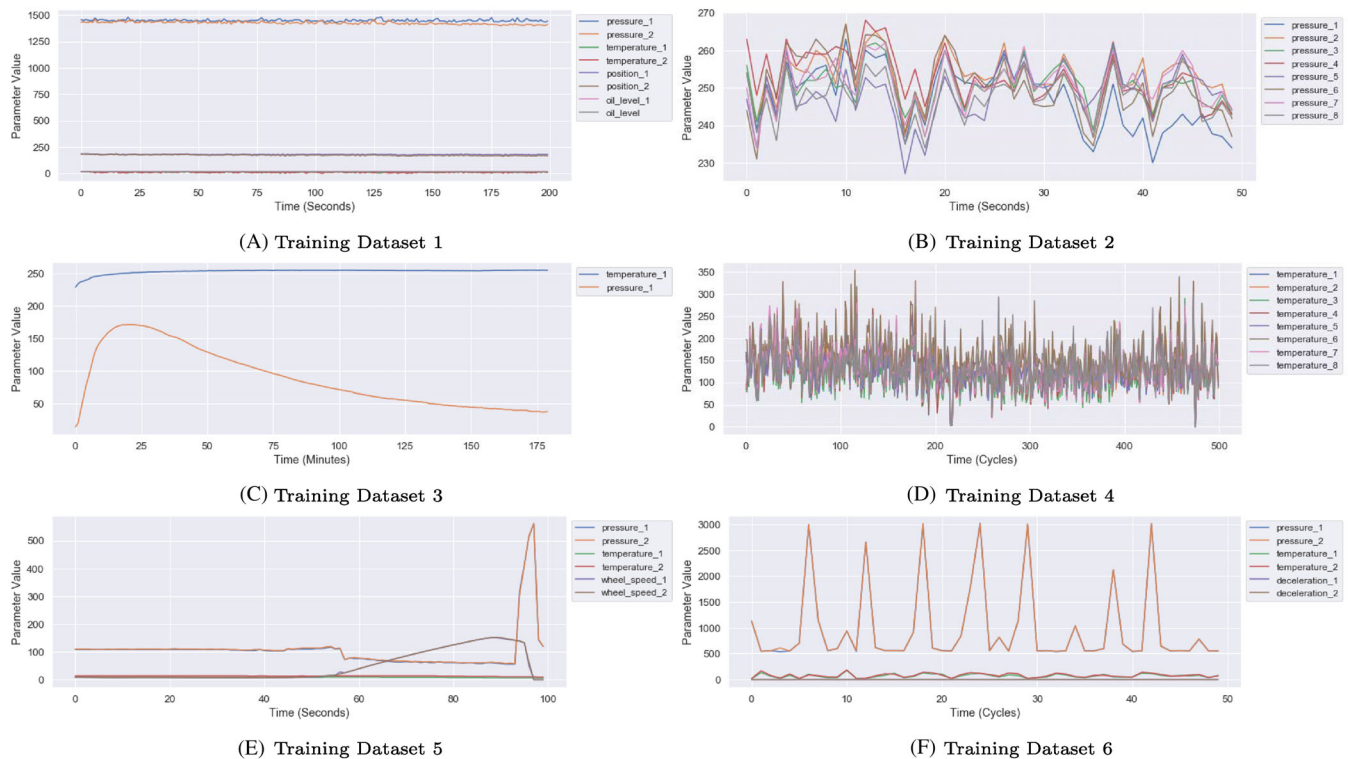


FIGURE 2 Typical example from the training datasets for each aircraft landing gear system.



and takeoff. This is a multivariate time series dataset for a tyre pressure indication system. The datasets consist of pressure readings from 8 sets of tyres from a set of Airbus aircraft after landing, and track the decline in tyre pressure due to leakage. The dataset consists of 40 sets of data, each 200 flight cycles long, across 8 pressure parameters. Like with the Hydraulics dataset, regression is used to estimate the RUL of each tyre, so they can be refilled. No changes were made to the original data, as these datasets were not long enough to cut down. These datasets contain a lot of noise from environmental factors, as ambient temperature and altitude can greatly impact the pressure in the tyres. Many of the datasets also contain a spike in pressure somewhere in the data where tyres have been refilled. An example of this data can be seen in Figure 2B.

### 4.3 | Training dataset 3: Tyre pressure indicator

Aircraft brakes play a pivotal role in ensuring safe landings and deceleration during the landing flight phase, with brake temperatures reaching as high as 800°C during intense braking. Regular maintenance is imperative to prevent potential malfunctions that can lead to critical issues, such as decreased stopping power or even brake failure. This multivariate dataset records brake temperature and tyre pressure for an Airbus aircraft as it lands. By classifying whether the rise and fall in brake temperature of a set of data is normal or abnormal, damaged or failed brakes can be identified. The dataset consists of 9283 sets of data, each 180 time steps long with each step representing a minute, with two parameters, brake temperature and tyre pressure. The normality of the data can be classified using unsupervised models such as Autoencoders. The original raw data consisted of a 9283-row, 360-column dataset where each column represented a brake or pressure reading. This was reshaped to a suitable size, and the sets of data have not been cut down or normalised. Each set of data shows the brake temperature sharply increases as the aircraft brakes and slowly decreases after the peak is reached, with a small uptake in pressure. An example of this data can be seen in Figure 2C.

### 4.4 | Training dataset 4: Landing gear brakes

Aircraft brake systems consist of intricate internal components like bearings, pads, and hydraulic mechanisms. These elements endure substantial mechanical stress during braking, making their maintenance crucial for optimal performance. This multivariate time series dataset contains the brake temperatures for eight sets of brakes. Anomalies in brake temperature or deceleration may not always indicate a fault with bearings, but when both are anomalous it indicates the bearings are damaged and need replacing. This dataset can be used to train classification models to differentiate between normal and anomalous readings, however, there is very little failure data available due to the rarity of its damaged bearing being identified in the field. The dataset consists of 194 sets of data, each 500 flight cycles long, across eight parameters for brake temperature. Clustering models like the K-Means clustering algorithm can be used to plot the mean brake temperature to identify anomalous readings. Each set of data appears as a linear dataset suffering from a lot of noise. An example of this data can be seen in Figure 2D.

### 4.5 | Training dataset 5: Landing gear brakes

There is a physical and mechanical relationship between the brake temperature, pressure and wheel for each wheel on the aircraft. Spikes in brake pressure will align with a sharp decrease in wheel speed as the brakes are applied, and by observing the correlation between these three variables can be used to identify anomalies. Sometimes you can only identify these anomalies when you have all the data together. This multivariate dataset records brake temperature, brake pressure and wheel speed for an Airbus aircraft during the last 100 of takeoff. Currently, it is being used to calculate deceleration to identify anomalies, which is reflected in Section 4.6. The dataset consists of 109 sets of data, each 100-time steps long with each step representing a second during the landing phase. It was filtered so the 80th point in each dataset was the max wheel speed to ensure a suitable capture of the wheel speed before and after was recorded. There is a unique ID for each landing, a date/time stamp, as well as temperature, pressure, and wheel speed values for each of the 8 wheels on the aircraft, with a total of 26 parameters. The original raw data consisted of a 58,000-row, 360-column dataset with several identifier parameters which have been removed for anonymisation. Each set of data ends in a spike of brake temperature as the wheels hit the tarmac, followed by a small spike in pressure indicating the brake has been applied which causes the wheel speed to decrease. An example of this data can be seen in Figure 2E.

## 4.6 | Training dataset 6: Landing gear brakes

Some components, such as wheel bearings in aircraft landing gear wheels, are difficult to ascertain the health of directly as they are inaccessible and direct measurements are difficult to take. However, an elevated max temperature and abnormal deceleration during landing can be indicative of their degradation during landing, allowing for specific wheels or brakes to be maintained post-landing. This multivariate dataset records the max brake temperature, max brake pressure, max wheel speed and deceleration for each wheel on an Airbus aircraft as it lands. This dataset is essentially the calculated values of the datasets from Section 4.5, where the max values for each parameter were identified, and the deceleration. The deceleration was calculated using the following equation, where  $a$  is the acceleration,  $v_f$  is the wheel speed at its peak,  $v_i$  is the wheel speed when the brakes are applied (spike in brake pressure detected), and  $t$  is the time taken between the peak wheel speed and the brake being applied.

$$a = \frac{v_f - v_i}{t} \quad (2)$$

The dataset consists of 37 sets of data, each 50-time steps long with each representing a flight cycle. There is a unique ID for each landing, a date/time stamp taken from the first row of data for each respective landing, eight parameters each for max temperature, pressure, wheel speed and calculated deceleration values, with a total of 34 parameters. Due to the similar structure to Section 4.6, clustering models can also be used here to plot the parameters against the deceleration values, and each set of data appears as a set of linear datasets suffering from noise. An example of this data can be seen in Figure 2F.

## 5 | RESULTS AND EVALUATION

Using the described methodology, DoppleGANger Models were trained for each of the six training sets. The hyperparameters of the models were tuned to maximise the performance of the selected metrics, quantitatively by minimising the KL divergence and TSTR error, and then with visual inspection using the qualitative methods. Using these tuned models, datasets containing 200 separate sets of data were generated. Typical examples of showing all the features for each of the new synthetic datasets generated from these models are displayed in Figure 3. The datasets are publicly available on the UWE Bristol Research Data Repository.<sup>30</sup>

### 5.1 | Autocorrelation

A qualitative evaluation of the autocorrelation was performed against the synthetic data. Firstly, the two datasets with the least MSE between them were selected based on the assumption that these two datasets will have the closest shape for comparison. The synthesised data aligns with the original data well in these cases, and the autocorrelation for each of these dataset pairs was calculated and displayed alongside the original data comparison in Figure 4. Note that for Figure 4E,F, only six out of 24 of the parameters are displayed to make it easier to interpret. Across all six selected datasets, the autocorrelation of the synthetic data aligns with the original across most of the parameters but contains enough diversity to not directly copy any of the original data. This result qualitatively confirms their statistical similarity. This consistency in shape across different datasets indicates a common underlying pattern or trend in their temporal behaviour.

### 5.2 | Pearson correlation heatmap

The Pearson correlation coefficient was calculated between each feature against every other feature to discern any disparities in the correlation between various features. This was done for the original datasets and the synthetic dataset, and then an absolute difference heatmap could be calculated by subtracting one from the other. The aim was to minimise the difference between the correlation connection of the original to the synthetic and achieve a difference value as close to zero as possible, with no values rising above 0.2 to account for some diversity within the datasets. The results of this analysis are presented in Figure 5 for each dataset. Most of the results are highly accurate, with the mean coefficient difference

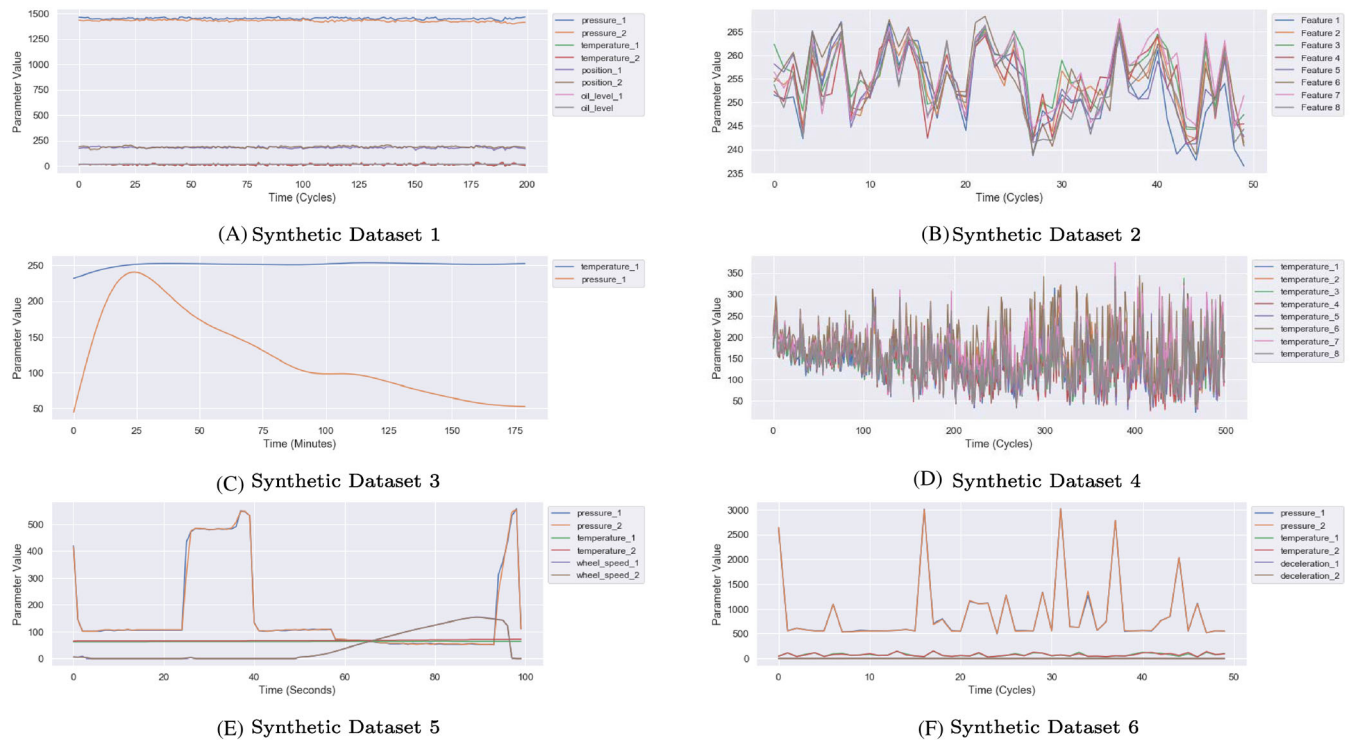


FIGURE 3 Typical synthesised data for each dataset.

values never approaching 0.1, or even exceeding 0.057 as shown in Table 2. There is a noticeable rise in this difference value for feature 2 in Synthetic Dataset 1 indicating that the feature specifically was not captured as well. Despite this, this final dataset scored the best across all metrics, and the values are still below 0.2, so it is acceptable. Despite this exception, these findings demonstrate the overall effectiveness of capturing the desired relationship and shape characteristics within the analysed features.

### 5.3 | Principal component analysis

An analysis of the six sets of data was conducted using PCA. The PCA across the random selection of test datasets is shown in Figure 6. Due to the presence of noise and the linearity of the data, no specific shape could be discerned by the PCA analysis for Synthetic datasets 2 and 4, however, it is important to note that the original and synthetic points are still aligned in the plot, suggesting the existence of a shared underlying structure. All six synthetic datasets exhibited similar patterns to the original training datasets in their respective PCA plots. While the specific shape might not be evident, the alignment suggests a consistent representation of shared information. This finding underscores the robustness of the synthesised data in capturing essential features of the original dataset, further substantiating the fidelity of the synthetic data generation approach.

### 5.4 | Kullback Leibler

The training data and the random selection of synthetic data were converted into discrete probability distributions. The KL divergence between these distributions was calculated and recorded, with the average and standard deviation of each listed in Table 3. Experimentation was conducted by comparing the training data to a noisy copy of itself, and it was found that most datasets were still recognisable with uniformly distributed noise of size 10% of the maximum value in the dataset, and a KL reading of less than 0.2. Minimising the KL score was one of the goals for tuning the hyperparameters using Optuna as described in Section 3.6. The KL score for the training data against a small random sample of itself was also calculated, to measure the diversity between the individual datasets in the training data. Both sets of calculated values

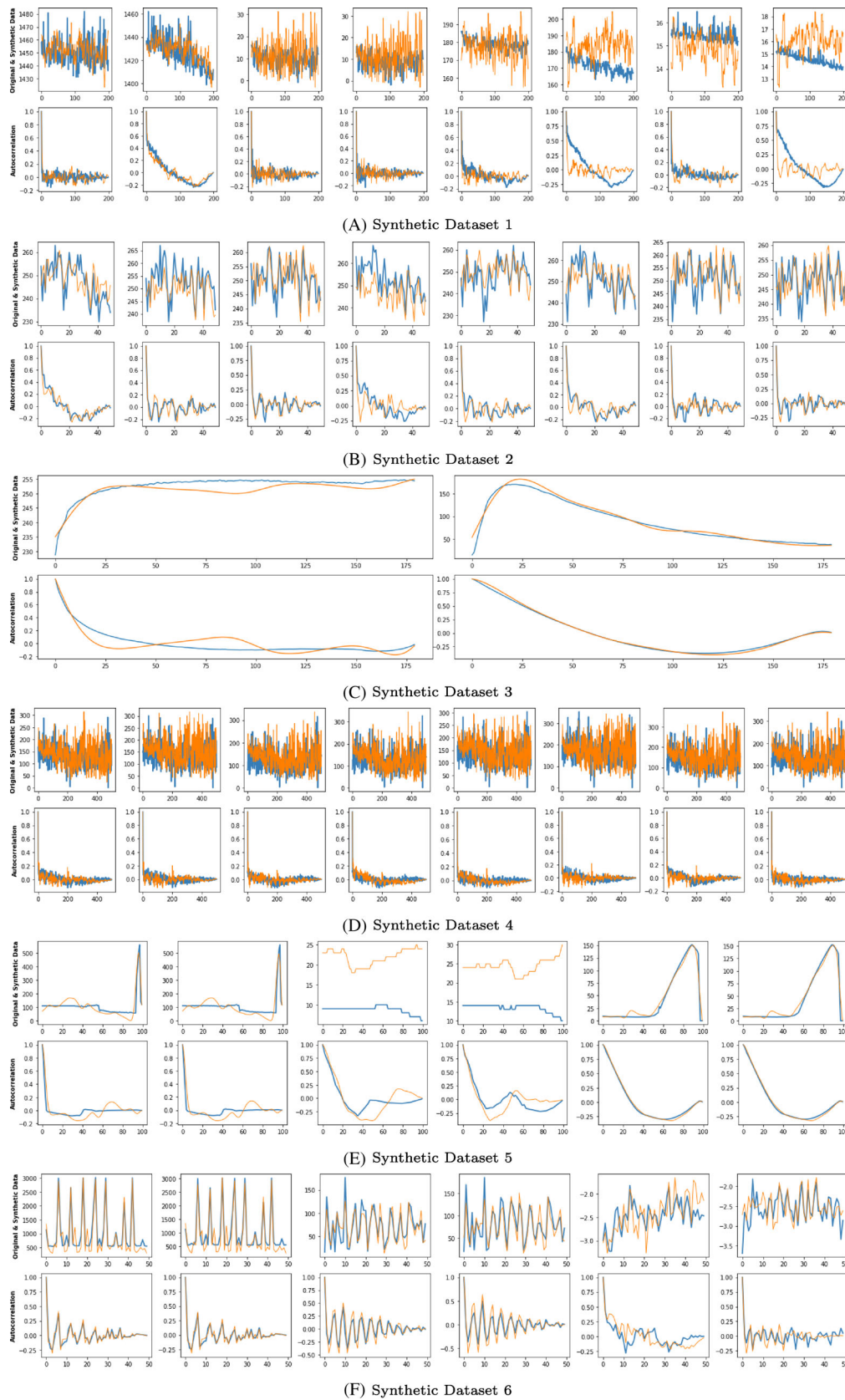


FIGURE 4 The autocorrelation of the mean features of typical data set from.

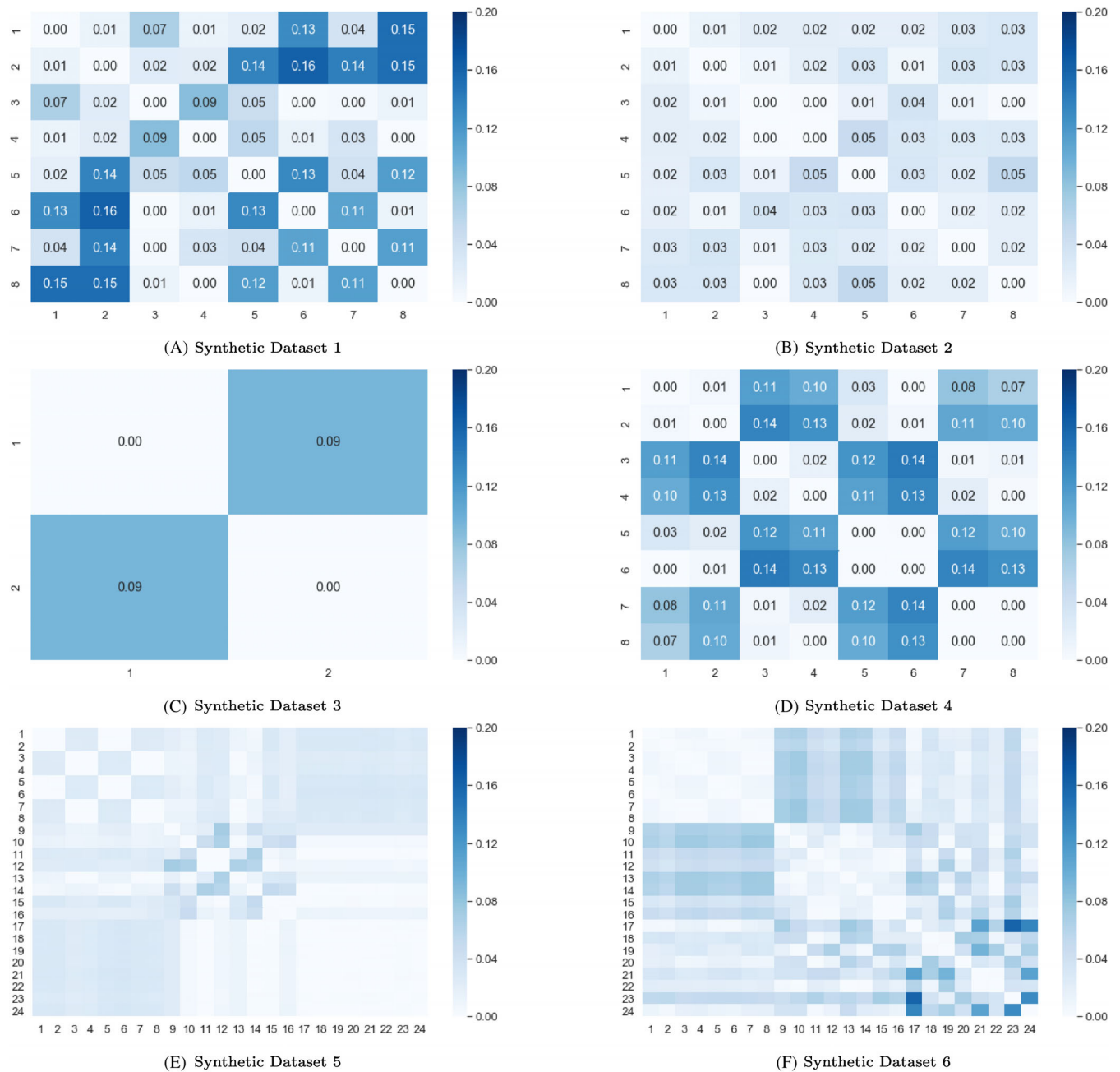


FIGURE 5 Difference in Pearson correlation coefficient heat maps for each synthetic dataset.

TABLE 2 The correlation mean difference for each dataset.

Dataset identifier	Mean correlation difference
Synthetic Dataset 1	0.057
Synthetic Dataset 2	0.021
Synthetic Dataset 3	0.056
Synthetic Dataset 4	0.062
Synthetic Dataset 5	0.017
Synthetic Dataset 6	0.029

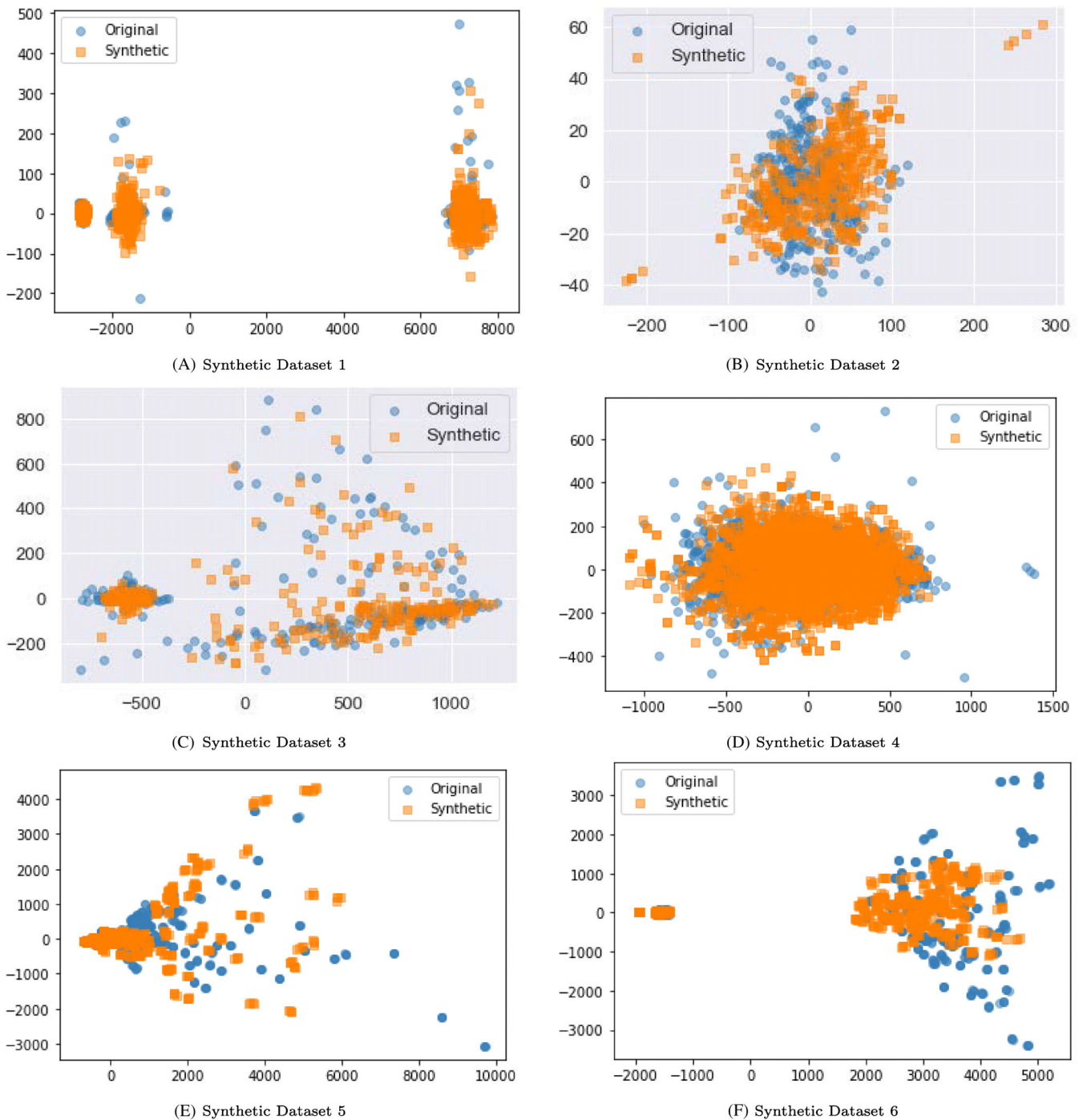


FIGURE 6 The principle component analysis graphs for each of the training and synthetic datasets.

are shown in Table 3. The average mean KL score for each of the datasets stays in the region of 0.2, and the difference in mean KL score for the synthetic and training values never exceeds 0.1, providing a quantitative assurance that even a randomly selected set of data is similar to the original.

## 5.5 | Train on synthetic, test on real

A pair of TensorFlow Sequential LSTM models were created and trained using the training data and randomly selected synthetic data perspective, using 25% of each for validation testing as they trained. These trained models were both

**TABLE 3** The mean and standard deviation of KL divergence for each dataset (S. for synthetic datasets, T. for training datasets).

Dataset identifier	S. Mean KL	S. Std KL	T. Mean KL	T. Std KL
Synthetic Dataset 1	0.232	0.230	0.301	0.198
Synthetic Dataset 2	0.215	0.125	0.265	0.150
Synthetic Dataset 3	0.321	0.164	0.389	0.346
Synthetic Dataset 4	0.267	0.239	0.222	0.274
Synthetic Dataset 5	0.504	0.508	0.530	0.527
Synthetic Dataset 6	0.241	0.161	0.194	0.203

**TABLE 4** The performance of an RNN for TRTR and TSTR tests.

Dataset identifier	TRTR MAE	TSTR MAE	Difference
Synthetic Dataset 1	0.323	0.169	-0.154
Synthetic Dataset 2	0.332	0.402	0.070
Synthetic Dataset 3	0.111	0.189	0.078
Synthetic Dataset 4	0.293	0.301	0.008
Synthetic Dataset 5	0.155	0.192	0.037
Synthetic Dataset 6	0.300	0.341	0.041

**TABLE 5** The reproduction accuracy of the original and synthetic datasets.

Landing gear system	Original	Synthetic	Difference
Synthetic Dataset 1	11.575%	20.046%	8.471%
Synthetic Dataset 2	18.709%	26.268%	7.559%
Synthetic Dataset 3	16.512%	20.004%	3.492%
Synthetic Dataset 4	27.034%	34.684%	7.650%
Synthetic Dataset 5	22.101%	23.049%	0.948%
Synthetic Dataset 6	26.838%	28.595%	1.757%

tested against the training data, predicting the next step in the series, and the MAE was recorded. The average MAE for each dataset is shown in Table 4. In each case, the performance metric of the synthetic data was slightly worse, however, they are within a comfortable threshold with no synthetic dataset having a higher MAE difference than 0.078.

### 5.5.1 | Time series classification

An LSTM Autoencoder was created using a TensorFlow Sequential model, to reconstruct the input and find the percentage difference between the original and reconstructed data. For each dataset, the model was trained on a randomly selected 75% of the original data, and tested against the remaining 25%. From comparing the performance of the model training of the training data to test data from the original, a prediction error of 20% was found to be a good indicator of similarity to the original dataset, The synthetic data was then passed into the model, and the prediction error. The prediction errors for the original and prediction data are shown in Table 5. For all six datasets, the prediction error for the synthetic data is just above, or even above the original, giving a strong indication of similarity. This method could potentially be used in the future to trim the dataset of low percentage error sets of data, to improve the overall accuracy.

## 5.6 | Applications for synthetic datasets

This data can now be used to train models for PdM problems.

### 5.6.1 | Synthetic datasets 1 and 2

The original purpose of the training datasets for Synthetic Datasets 1 and 2 was RUL estimation for the point at which the data crossed an established failure threshold. While the real threshold for failure cannot be shared, an artificially selected value could be used to build models. The data ranges between 1345.87 and 1544.97, so a suitable value for a failure threshold would be 1400 for example. A model can now be generated to either predict how many steps are remaining until this point is crossed, or whether the failure point is within the next 20 time steps. This could be performed using a basic linear regression model, or ARIMA.

### 5.6.2 | Synthetic dataset 3

The friction from landing heats aircraft landing gear tyres up, which spikes the tyre pressure for the period after landing. Technicians performing maintenance need to record the standard tyre pressure once the temperature drops and steadies, however, this process of cooling can take 3 h. So using only the first 5–10 data points, can an estimation of all of the remaining data points, particularly the last one, be estimated? No changes or assumptions are required to build the same type of model for this synthetic data.

### 5.6.3 | Synthetic datasets 4 and 6

The original motivation for tracking the peak brake temperature and deceleration during takeoff was to identify faults with the wheel bearings. The health of wheel bearings can be difficult to directly measure, so for instances when they do fail, they have been linked to possible indicators such as anomalous brake and deceleration values. For the original data, after a thorough analysis, an equation was written to indicate the regions of both high brake temperature and deceleration which indicated bearing failure. Without known examples due to the too few examples of failure to synthesise data, an unsupervised method such as k-means clustering or support vector machines can be used to identify a region of “normal behaviour.”

### 5.6.4 | Synthetic dataset 5

The raw data for the takeoff wheel speed, brake pressure and brake temperature can be used to identify landing gear faults, even without calculating the max and deceleration values. While no examples of failures in the training data were identified, unsupervised classification models such as an LSTM autoencoder can be trained on only “normal” data. The reconstruction accuracy can be used to indicate “normal” data and anomalous data, indicating faults. In this instance, artificially anomalous data (e.g., Normal example but with noise, flat increase in values etc.) will need to be used to test the model. This method can just as easily be applied to the synthetic dataset.

## 6 | CONCLUSION

The fidelity testing indicates that the synthetic data produced in this study resembles the original Airbus datasets in a way that could be used by researchers and could be used for testing existing and novel PdM methods. This demonstrates the feasibility and effectiveness of Gretel.ai as a tool for data augmentation for industry data. The six new multivariate datasets have been made available to the public from the University of the West of England library service<sup>30</sup> for researchers to test novel PdM methods against. The names for each synthetic dataset can be seen in Table 6. This increases the diversity of available datasets and can be used to improve the size of training data for improved performance of existing methods,



TABLE 6 The file names for each of the synthetic datasets.

Dataset identifier	File name
Synthetic Dataset 1	BogiePitchTrimmer_PressureDrop_OverFlightCycles
Synthetic Dataset 2	TyrePressureIndicator_PressureDrop_OverFlightCycle
Synthetic Dataset 3	TyrePressureIndicator_TemperatureRise_DuringLandin
Synthetic Dataset 4	LandingGearBrakes_MaxTemperature_OverFlightCycles
Synthetic Dataset 5	LandingGearBrakes_Deceleration_DuringTakeof
Synthetic Dataset 6	LandingGearBrakes_Deceleration_OverFlightCycles

both within the industry and academic settings. The methodology presented in this paper answers all of the proposed research questions and can be used as a framework for generating data in other industries, increasing the diversity of available data to PdM researchers.

There is further work to be done in this field, exploring a greater diversity of aircraft systems and using these new public datasets to compare novel models and frameworks. The landing gear system was selected due to its vital importance for the operation of the aircraft, and the lack of existing public data. There are however many additional systems that have been identified that lack existing public datasets, including pneumatics, fuselage and power units.<sup>7</sup> More Airbus datasets were available, and synthetic facsimiles for these could be released alongside this data in the future, but this paper focuses on selecting and fine-tuning models to achieve high performances. The Validation metrics used for confirming the similarity between the original and synthetic data were not intended to prove that data augmentation would improve the performance of the original methods. In future work, these datasets will be combined with the original and tested against the standard models being used by Airbus. The results of this study provide valuable insights into the potential of Gretel.ai and other listed models for expanding the available open data in this field and improving predictive maintenance for aircraft and other industries.

## AUTHOR CONTRIBUTIONS

**Izaak Stanton:** writing – original draft; methodology; software; visualization; writing – review and editing; conceptualization. **Kamran Munir:** supervision; writing – review and editing. **Ahsan Ikram:** supervision; writing – review and editing. **Murad El-Bakry:** supervision; writing – review and editing; resources.

## ACKNOWLEDGMENTS

This work has been jointly funded by the University of the West of England and Airbus under the UWE Project ID 6263917 Finance Project Code RCSC0112. Mr. Izaak Stanton is the PhD scholar funded by the project, Professor Dr Kamran Munir is the project lead, Dr Ahsan Ikram is the co-supervisor and Dr Murad El-Bakry is the industry lead from Airbus.

## CONFLICT OF INTEREST STATEMENT

There is no conflict of interest relevant to this article from any of the Authors.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/eng2.12946>.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in UWE Research Repository at <https://researchdata.uwe.ac.uk/717>.

## ENDNOTES

\*<https://researchdata.uwe.ac.uk/id/eprint/717/>

†<https://github.com/ydataai/ydata-synthetic>.

‡<https://github.com/gretelai/gretel-synthetics>.

## ORCID

Izaak Stanton  <https://orcid.org/0000-0003-0296-1780>

## REFERENCES

1. AIRBUS. Accidents by flight phase—Airbus Accident Statistics. 2018. <https://accidentstats.airbus.com/statistics/accident-by-flight-phase>
2. Hsu T-H, Chang Y-J, Hsu H-K, Chen T-T, Hwang P-W. Predicting the remaining useful life of landing gear with prognostics and health management (phm). *Aerospace*. 2022;9(8):462.
3. IATA's Maintenance Cost Technical Group. *Airline Maintenance Cost Executive Commentary Edition 2019*. International Air Transport Association: Technical report; 2019. <https://www.iata.org/contentassets/bf8ca67c8bcd4358b3d004b0d6d0916f/mctg-fy2018-report-public.pdf>
4. Okoro OC, Zaliskyi M, Dmytriiev S, Solomentsev O, Sribna O. Optimization of maintenance task interval of aircraft systems. *Comput Netw Inf Secur*. 2022;9(8):77-89.
5. NASA. Commercial Modular Aero-Propulsion System Simulation (C-MAPSS), Version 2(LEW-18315-2)—NASA Software Catalog. 2015. <https://software.nasa.gov/software/LEW-18315-2>
6. Xiong J, Fink O, Zhou J, Ma Y. Controlled physics-informed data generation for deep learning-based remaining useful life prediction under unseen operation conditions. *Mech Syst Signal Process*. 2023;197:110359. doi:10.1016/J.YMSSP.2023.110359
7. Stanton I, Munir K, Ikram A, El-Bakry M. Predictive maintenance analytics and implementation for aircraft: challenges and opportunities. *Syst Eng*. 2023;26(2):216-237. doi:10.1002/SYS.21651
8. Parthasarathy D, Backstrom K, Henriksson J, Einarsdottir S. Controlled time series generation for automotive software-in-the-loop testing using GANs. 2020 IEEE International Conference on Artificial Intelligence Testing, AITest 2020, Oxford, UK, Institute of Electrical and Electronics Engineers Inc. 2020;39-46. doi:10.1109/AITEST49225.2020.00013
9. Xinyu G, See KW, Liu Y, Arshad B, Zhao L, Wang Y. A time-series Wasserstein GAN method for state-of-charge estimation of lithium-ion batteries. *J Power Sources*. 2023;581:233472. doi:10.1016/j.jpowsour.2023.233472
10. Long Y, Liu L, Shao L, Shen F, Ding G, Han J. From zero-shot learning to conventional supervised classification: unseen visual data synthesis. Proceedings 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, Hawaii, Institute of Electrical and Electronics Engineers Inc. 2017;2017:6165-6174. doi:10.1109/CVPR.2017.653
11. Rankin D, Black M, Bond R, Wallace J, Mulvenna M, Epelde G. Reliability of supervised machine learning using synthetic data in health care: model to preserve privacy for data sharing. *JMIR Med Inform*. 2020;8(7):e18010. doi:10.2196/18910
12. Dahmen J, Cook D. SynSys: a synthetic data generation system for healthcare applications. *Sensors*. 2019;19(5):1181. doi:10.3390/S19051181
13. Takahashi S, Chen Y, Tanaka-Ishii K. Modeling financial time-series with generative adversarial networks. *Phys A: Stat Mech Appl*. 2019;527:121261. doi:10.1016/J.PHYSA.2019.121261
14. Boutros F, Huber M, Siebke P, Rieber T, Damer N. SFace: privacy-friendly and accurate face recognition using synthetic data. 2022 IEEE International Joint Conference on Biometrics, IJCB 2022, Pages 1–11, Abu Dhabi, United Arab Emirates, Institute of Electrical and Electronics Engineers Inc. 2022. doi:10.1109/IJCB54206.2022.10007961
15. Wang Q, Gao J, Lin W, Yuan Y. Learning from synthetic data for crowd counting in the wild. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, IEEE Computer Society. 2019;2019:8190-8199. doi:10.1109/CVPR.2019.00839
16. Barisic A, Petric F, Bogdan S. Sim2Air - synthetic aerial dataset for UAV monitoring. *IEEE Robot Autom Lett*. 2022;7(2):3757-3764. doi:10.1109/LRA.2022.3147337
17. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Inf Process Syst*. 2014;27:2672-2680. <http://www.github.com/goodfeli/adversarial>
18. Tianlin X, Wenliang LK, Google MM, Acciaio B. COT-GAN: generating sequential data via causal optimal transport. *Adv Neural Inf Process Syst*. 2020;33:8798-8809. <https://proceedings.neurips.cc/paper/2020/hash/641d77dd5271fca28764612a028d9c8e-Abstract.html>
19. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. International Conference on Learning Representations 2016, San Juan, Puerto Rico. 2015 arXiv:1511.06434.
20. Zhang H, Xu T, Li H, et al. StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2016 5907-5915.
21. Brophy E, Wang Z, She Q, Ward T. Generative adversarial networks in time series: a systematic literature review. *ACM Comput Surv*. 2023;55(10):31. doi:10.1145/3559540
22. Esteban C, Hyland SL, Rättsch G. Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs. arXiv preprint, arXiv:1706.02633 2017.
23. Deng L. The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Process Mag*. 2012;29(6):141-142. doi:10.1109/MSP.2012.2211477
24. Yoon J, Jarrett D, van der Schaar M. Time-series generative adversarial networks. *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*. 2019;32:5508-5518.
25. Lin Z, Jain A, Wang C, Fanti G, Sekar V. Using GANs for sharing networked time series data: challenges, initial promise, and open questions. Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC, Association for Computing Machinery. 2020;464-483. doi:10.1145/3419394.3423643
26. Wang S, Rudolph C, Nepal S, Grobler M, Chen S. PART-GAN: privacy-preserving time-series sharing. Artificial Neural Networks and Machine Learning—ICANN 2020: 29th International Conference on Artificial Neural Networks, Bratislava, Slovakia, Springer Science and Business Media Deutschland GmbH. 2020;12396 LNCS:578-593. doi:10.1007/978-3-030-61609-0\_46/COVER

27. SKYbrary. Aircraft Condition Monitoring System (ACMS)—SKYbrary Aviation Safety. 2021. <https://www.skybrary.aero/articles/aircraft-condition-monitoring-system-acms>
28. SKYbrary. Meteorological Aerodrome Report (METAR) — SKYbrary Aviation Safety. 2021. <https://skybrary.aero/articles/meteorological-aerodrome-report-metar>
29. Optuna. Optuna-A hyperparameter optimization framework. 2023. <https://optuna.org/>
30. Stanton I, Munir K, Ikram A, Elbakry M. Aircraft Predictive Maintenance Automation and Optimisation: Synthesised Landing Gear Datasets. 2023. <http://researchdata.uwe.ac.uk/717>

**How to cite this article:** Stanton I, Munir K, Ikram A, El-Bakry M. Data augmentation for predictive maintenance: Synthesising aircraft landing gear datasets. *Engineering Reports*. 2024;e12946. doi: 10.1002/eng2.12946