

KEY PARAMETER IDENTIFICATION IN AIRCRAFT STRAIN PREDICTION MODELS

Kate Niessen[1] & Daniel Franke[1]

[1]Defence Science and Technology Group, 506 Lorimer Street, Fishermans Bend, Victoria, Australia, 3207

Abstract

In order to prevent the catastrophic failure of aircraft structures due to fatigue, modern airworthiness management practices require operational usage to be tracked at both a fleet and individual aircraft level. These individual aircraft tracking programs (IATP) can vary in fidelity, with the most basic systems monitoring only key events (take-offs, landings, pressure-cycles etc.) and the most sophisticated systems employing networks of strain gauge sensors installed at critical airframe locations on every aircraft across a fleet. In practice, these high fidelity IATP systems are not widely used, as they require ongoing expensive fleet-wide maintenance and periodic strain sensor calibration. As a result, intermediate fidelity IATPs that rely on recorded flight parameter time histories, are a commonly implemented approach.

Recent advances in artificial intelligence (AI) provide an opportunity to enhance intermediate fidelity IATPs through the prediction of airframe strain at fatigue critical locations using just flight parameter data. Parameter-strain models can be trained using data from a single instrumented flight test aircraft and subsequently applied across a fleet where only parameter data is captured. A key challenge in this approach is the identification of the most influential flight parameters from a large candidate pool, ensuring accurate strain prediction while maintaining computational efficiency and model interpretability.

This paper investigates and compares two parameter selection strategies using artificial neural network (ANN) based strain prediction models. The first strategy employs an exhaustive approach, generating all possible combinations of ANN input parameters and selecting the 'best' model based on minimised root mean square error (RMSE) between measured and predicted strain. The second strategy uses Shapley additive explanations (SHAP) values to rank input parameters by relative importance, allowing reduced-order models to be developed from a single trained ANN.

Results show that the exhaustive approach achieved superior prediction accuracy with intuitive input parameters, while the SHAP-based models achieved comparable performance for moderate model sizes with greatly reduced analysis time. These findings demonstrate that SHAP-based parameter selection offers a practical and efficient pathway for rapid development of AI-enhanced intermediate fidelity IATPs, while exhaustive methods remain advantageous when maximum accuracy is required.

Keywords: parameter strain models, neural networks, individual aircraft tracking system

1. Introduction

The prevention of fatigue failures in aircraft structures is a fundamental objective of airworthiness management programs. Fatigue damage accumulates as a function of both load magnitude and repetition and is driven by the usage of each individual aircraft. To address this requirement, modern fleets often employ IATPs which monitor usage and estimate fatigue consumption throughout the service life of the aircraft.

As high fidelity IATP systems can be cost prohibitive in terms of installation, maintenance and calibration requirements, intermediate fidelity systems based on flight parameter recordings are

commonly deployed across both military and civilian aircraft fleets. Typically, most modern aircraft monitor and record a broad suite of flight parameters, offering a cost-effective data source for usage tracking. These systems are popular as they can also be used to capture data which enables condition based monitoring/maintenance and flight operation quality assurance. From a fatigue monitoring perspective however, the relationship between parameters and structural strain and load is indirect. Consequently models are required which infer structural response based on aircraft configuration, flight/ground conditions and mass parameters.

Recent advances in data-driven modelling and artificial intelligence have enabled the development of parametric strain prediction models capable of estimating structural strain at fatigue-critical locations using routinely recorded flight parameters alone. Under this paradigm, a single heavily instrumented aircraft – typically within a flight loads test program (FLTP) – can be used to train models that relate measured strain to flight parameters. Once validated, these models can be deployed across a fleet of non-instrumented aircraft, enhancing IATP fidelity without any additional hardware. Computationally efficient techniques to identify key flight parameters are required to ensure that strain prediction models are of acceptable quality.

Understanding the importance of input variables in any mathematical modelling is crucial to the interpretation of modelled results. Modern aircraft can record hundreds of parameters, many of which are highly correlated or indirectly related to loads at a specific location. Incorporating unnecessary parameters increases computational expense, reduces model interpretability, and increase the risk of overfitting. Conversely, omission of key parameters degrades predictive accuracy. The ability to identify key input parameters from a wider pool based on the strength of their relationship with a target output (in this case strain) is vital in both reducing the input data space and interpreting individual parameter contributions in the final model. With simple models such as linear regression, the contribution of each input parameter is easier to assess, however with more complex modelling tools such as ANNs, the interpretation of relative variable importance becomes increasingly difficult. Balancing the accuracy of a complicated model against the interpretability of a simple model is challenging.

This study addresses the problem of flight parameter selection for ANN-based aircraft strain prediction in intermediate-fidelity IATP applications. Using data from a comprehensive FLTP conducted on the Leonardo C-27J (Spartan) aircraft, two contrasting parameter selection strategies are examined. The first employs an exhaustive search across all possible combinations of available inputs to identify the most accurate model. This method is computationally expensive as it requires all ANN model combinations to be trained in order for their numeric accuracy to be assessed. The second applies a more computationally efficient parameter ranking technique based on Shapley additive explanations and requires only a single ANN model to be trained. By comparing the predictive performance, computational requirements, and deployment (generalisation) behaviour of models derived from each approach, this work seeks to inform the development of robust and practical parametric strain models for IATP applications.

2. Background

The concept of indirect aircraft structural usage monitoring using operational parameters has been explored extensively in the literature. Early work demonstrated that regression-based models could be used to infer loads from limited sets of flight parameters, particularly when guided by physical insight into the dominant contributors to structural response. However, such linear approaches struggle to represent nonlinear effects arising from things such as; control surface deflections, fuel distributions, manoeuvre severity and configuration changes.

Neural network methods were introduced to address these limitations, allowing strain or load responses to be learned directly from data without explicit specification of functional form [1]. Subsequent studies demonstrated that ANN-based models could reproduce measured strain time histories with high fidelity, provided that sufficient training data were available and that models were trained across an appropriate flight parameter envelope.

Despite these successes, most investigations have relied on expert judgement to select key parameters as network inputs. This approach is effective when supported by strong domain

knowledge, but it becomes increasingly difficult as the number of available parameters grows. Furthermore, expert-chosen inputs may reflect prevailing assumptions about load drivers that do not fully capture the complexity of the real structural response, particularly in modern aircraft with flexible structures and modern control systems. Secondary relationships between structural response and input parameters can be difficult to predict as the individual impact may be small but important.

Recent advances in model transparency offer new possibilities for parameter selection. Shapley-based methods take a unified approach to interpreting model predictions using analysis of multiple additive feature importance methods [2]. The Shapley framework makes use of game theory to determine the relative contribution of a particular feature at specified query points through the calculation of Shapley values. These values are determined through the use of a simplified 'explainer model' which is an approximation of the original more complex ANN. In the context of machine learning, Shapley values correspond to the deviation of the ANN prediction from the average prediction due to a particular feature or input variable for each query point. Therefore, for each query point, the sum of the Shapley values across all features will be equal to the total deviation of the ANN prediction from the average. For each feature, a single Shapley value may be calculated by averaging across the specified set of query points, giving the ability to rank input parameters based on their importance to the model. Shapley values are calculated after model training and are agnostic to the training algorithm used [3]. This approach has gained traction in domains where transparency and interpretability are critical, including aerospace applications.

3. Flight Loads Test Program and Data Set

As part of a collaborative program between the Commonwealth of Australia and Leonardo SpA a dedicated FLTP was undertaken in order to collect representative load and strain data on a Leonardo C-27J aircraft. The program employed a heavily instrumented aircraft equipped with 362 strain gauge bridges distributed across the wing, fuselage, empennage and nacelle structures. The FLTP was conducted primarily in Turin, Italy and was composed of two parts: a static load ground calibration conducted over the timeframe June 2020 to January 2021, and a flight test program conducted over the timeframe March to November 2021.

For the parameter to strain relationships developed in this study, a subset of five flights from the FLTP data set was examined (Table 1). Four of the flights consisted primarily of dedicated loads flight test manoeuvres, while one represented a nominal operational mission. These flights were chosen to cover a range of structural loading conditions and encompassed significant variations in speed, altitude, acceleration, fuel state and configuration. Analysis was based on the airborne sections of the flight only as identified by the weight on wheels parameter timings.

Table 1 – Selected flights for analysis

Flight	Duration (hrs)	Description
F023	2.1	flight test manoeuvres
F024	2.1	flight test manoeuvres
F062	1.3	flight test manoeuvres
F098	1.1	flight test manoeuvres
F107	2.5	mission profile

Of the total 362 gauges fitted to the test aircraft, eleven were identified as suitable candidates for parametric IATP modelling (Table 2). These gauges were felt to be the most representative of key airframe locations for fatigue monitoring as previously identified by Leonardo or were locations of general interest for tracking purposes [4]. For this study, a single wing gauge (WPR03AMC1) located at rib 3 (wing root) on the upper middle spar cap of the port wing was chosen for analysis (Figure 1). This location predominantly experiences axial strain driven by bending loads in the wing middle spar cap. Focusing on one gauge enabled rigorous exploration of parameter selection techniques.

Table 2 – Gauges available for analysis (selected gauge indicated by shading)

ID	Structure	Gauge name	Gauge type	Location description	Location match with Leonardo IATP location
W1	Wing	WPR03AMC2	Axial	Port wing - lower middle spar - lower spar cap (rib 3)	Close
W2	Wing	WPR03AMC1	Axial	Port wing - upper middle spar - upper spar cap (rib 3)	Close
W3	Wing	FPF18AUK1	Axial	Port – fuselage - wing attachment	Exact
F1	Fuselage	FPF09ALS1	Axial	Port – fuselage – crew door surround – skin	Close
F2	Fuselage	FPF30ALG1	Axial	Port – fuselage – ramp door latch	Similar
H1	Horizontal stabiliser	FPF42AUL1	Axial	Port – fuselage – horizontal stabiliser attachment – upper lug	Exact
V1	Vertical stabiliser	VPR02SRS1	Shear	Port – vertical stabiliser – skin panel – skin (rib 2a)	Similar
N1	Nacelle	NPR13ALL1	Axial	Port – nacelle – lower attachment	Exact
W4	Wing	WSR03AMC1	Axial	Starboard wing - upper middle spar - upper spar cap (rib 3)	n/a
W5	Wing	WSR03BFC1	Bending	Starboard wing - front spar - spar cap (rib 3)	n/a
W6	Wing	WPR13AMC2	Axial	Port wing - lower middle spar cap - lower spar cap (rib 13)	n/a

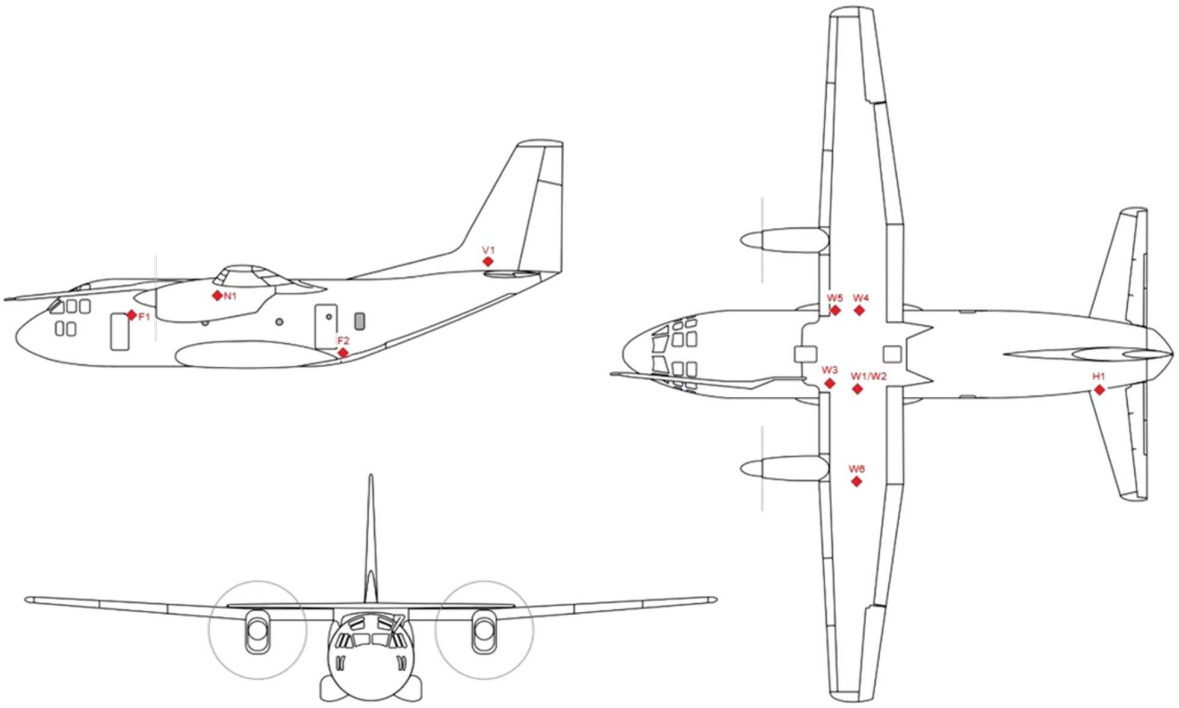


Figure 1 – Three way aircraft view showing gauge locations [5]

Prior to the analysis, the strain response (recorded at 512 Hz) was filtered using a 1 Hz low pass filter to remove any dynamic content[1].

A core requirement of parametric IATP modelling is that input parameters be available not only during flight testing but also during routine operations utilising non-instrumented fleet aircraft. Accordingly, the pool of input flight parameters was defined by cross-referencing the list of available parameters

[1] A digital filter with a passband cut-off frequency of 1 Hz and a stopband frequency of 1.5 Hz was used.

from the FLTP (test) aircraft against the list of parameters routinely recorded on the Royal Australian Air Force C-27J aircraft fleet for fatigue tracking. The resulting parameter set is composed of fifteen continuous variables describing aerodynamic state, aircraft configuration, attitude and environmental conditions as well as fuel distribution and flight state (Table 3). Discrete variables, such as weight on wheels, were excluded. Fuel tank weights, calculated from available fuel quantity data[2] are included to capture mass distribution effects. Gross aircraft weight was deliberately excluded because it is not recorded as a parameter by the fleet recording system. This omission is a known limitation and is revisited in the discussion of model deployment behaviour.

Table 3 – Available input parameter pool

Description (Flight Parameter)	Description (Flight Parameter)
Angle of attack (ALPHA)	Vertical acceleration (NZ)
Angle of sideslip (BETA)	Pressure altitude (PALT)
Calibrated airspeed (CAS)	Pitch angle (PITCHANGLE)
Cabin differential pressure (CDP)	Pitch rate (PITCHRATE)
Flap deflection (FLAP)	Roll angle (ROLLANGLE)
Main fuel tank weight – port (FUELMAINPtotal)	Static air temperature (SAT)
Main fuel tank weight – starboard (FUELMAINStotal)	Vertical velocity, Embedded GPS Inertial navigation unit (VZEGI)
Lateral acceleration (NY)	

The FLTP parameter data was resampled down to the fleet wide sample rates and then up-sampled to 512 Hz and synchronised in order to align with the FLTP strain gauge sample rate.

4. Modelling Framework

All strain prediction models were developed using feed-forward artificial neural networks with a single hidden layer and a 100 epoch training cycle. The Levenberg-Marquardt algorithm was employed with a 70/15/15 percentage data split respectively for training, validation and testing. Two network sizes (number of neurons) were considered during initial development, however the final analysis focuses on networks with ten hidden neurons, as these provided sufficient modelling capacity without excessive training complexity.

Flight F062 (Table 1) was selected as the primary training flight due to its high manoeuvre content and broad parametric ranges, making it well suited for capturing strain-parameter relationships relevant to accurate strain prediction. All parameter selection and reference models discussed herein are trained using F062 only. Other flights are used exclusively for independent deployment and validation.

Model performance was quantified using RMSE calculated between measured and predicted strain values. For comparison across different flights, RMSE was also expressed as a percentage of the observed strain range within each flight.

5. Parameter Selection Methodologies

5.1 Exhaustive parameter selection method

The exhaustive approach to parameter selection involved generating and training neural networks for every possible combination of the fifteen candidate inputs. For a given model size, defined by the number of input parameters, all unique combinations were evaluated. The model yielding the lowest prediction error was selected as the ‘best’ model for that size.

[2] Main and auxiliary raw fuel tank weights were measured on the avionics bus and then calibrated using a hybrid linear/cubic model to increase accuracy. Fixed values of unusable trapped fuel in each tank (below the pump level) were then added to the calibrated data to allow for accurate fuel quantity data to be used for inertial load calculations.

This approach provides a rigorous upper bound on achievable performance for each parameter set. However, the number of possible combinations grows exponentially with the number of inputs, leading to substantial computational cost to develop all models. Even when parallel processing[3] was used, complete evaluation of all parameter subsets required several hundred hours of computation for a single training flight (Figure 2).

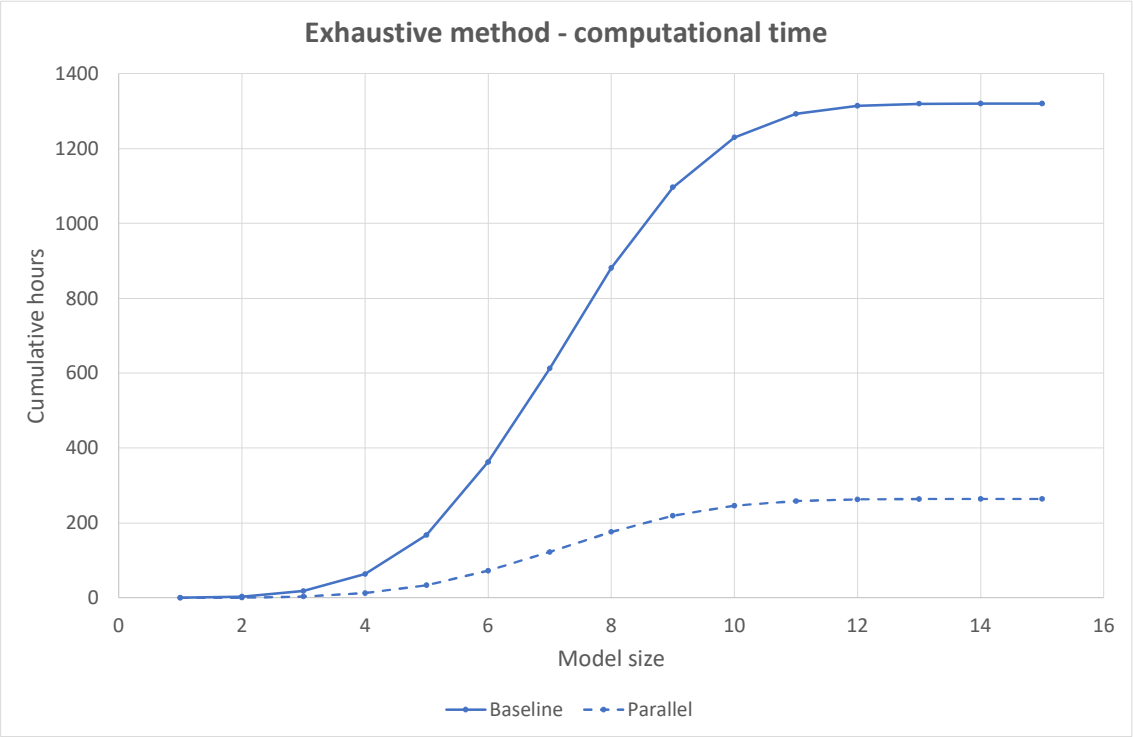


Figure 2 – Cumulative computing times (baseline and parallel processing), F062 training flight

Despite the computational expense, the exhaustive approach serves as a useful benchmark against which more efficient parameter selection methods can be evaluated.

5.2 Shapley based parameter ranking method

The Shapley-based method employed a single neural network trained using all fifteen flight parameters and using the same architecture as that used for the exhaustive approach. SHAP values were calculated using an interventional kernel algorithm with a maximum of 1024 subsets and a sampled observation of 100 (default). Values were calculated on a smaller subset of 'query points' which were randomly sampled from the wider data set. In this case, 0.5% of the data (~10,000 points) were used to approximate results from the entire set. Repeated calculations were performed to assess the stability of the resulting parameter rankings. Calculations showed good stability for the highest-ranked parameters, while moderate variability was observed among mid-ranked inputs. Importantly, overall model error for the full-input networks remained consistent across different modelling runs, indicating that ranking variability did not reflect instability in the underlying neural networks.

Parameters were ranked in order of importance according to their mean absolute SHAP values. Reduced-order models were then constructed by progressively adding parameters in order of ranked importance, producing an additive sequence of models analogous to forward stepwise regression.

[3] The baseline computational figures are taken from running the models using a desktop computer with an i7-9700 CPU at 3 GHz and 64 GB of RAM. Parallel processing (enabled by changes to MATLAB code) allows each of the 8 cores to act as an individual processor, with each core running a section of the central calculations simultaneously rather than in sequence, thereby reducing the overall computing time.

This approach dramatically reduced computational cost, requiring training of only a single network and subsequent post-processing to determine the Shapley values. In total, this process is completed in just sixty eight minutes (1.1 hours), and enabled rapid identification of influential parameters. However, it does constrain parameter selection to an additive framework which may result in models with lower accuracy being selected as compared to the exhaustive (best) method.

6. Results

6.1 Parameter identification

A comparison of the two parameter selection methods (best and Shapley), using F062 as the training flight, is given for a subset of results in Table 4.[4] Here, the Shapley model parameters are listed in order of decreasing importance and the %RMSE diagnostic is the RMSE as a percentage of the strain range for flight F062.

Table 4 – Wing root strain prediction - parameter selection and %RMSE

	Exhaustive (best) model		Shapley model	
nVars	**Parameters**	**%RMSE**	**Parameters**	**%RMSE**
2	NZ, ALPHA	1.38	CAS, FLAP	6.57
4	NZ, CAS, FLAP, NY	1.12	CAS, FLAP, ALPHA, NZ	1.18
6	NZ, ALPHA, CAS, NY, FUELMAINStotal, PITCHRATE	0.92	CAS, FLAP, ALPHA, NZ, PITCHRATE, FUELMAINPtotal	1.06
8	NZ, ALPHA, CAS, NY, FUELMAINStotal, PITCHRATE, FLAP, SAT	0.79	CAS, FLAP, ALPHA, NZ, PITCHRATE, FUELMAINPtotal, FUELMAINStotal, PALT	0.93

The exhaustive approach consistently identified vertical acceleration (NZ) as a dominant driver of wing-root spar cap axial strain, which aligns well with the expectation that wing root strain is proportional to total aircraft weight. Other consistently selected parameters from this approach include angle of attack (ALPHA), calibrated airspeed (CAS) and lateral acceleration (NY) accounting for both aerodynamic and inertial drivers. SHAP-based rankings identified CAS and flap deflection (FLAP) as highly influential parameters with ALPHA and NZ appearing in higher-order models.

Exhaustively selected models consistently outperformed SHAP-selected models of the same size with respect to the training data. This difference was most pronounced when fewer input parameters were used and where SHAP-based models omitted physically dominant parameters such as NZ in early selections. Once normal acceleration was incorporated, SHAP-based models more closely approached the accuracy of exhaustive selections.

For both parameter selection methods, prediction error decreased rapidly with increasing model size. Substantial improvements were observed when moving from two to four input parameters, after which performance gains diminished. Beyond approximately six input parameters, additional reductions in RMSE were considered marginal, indicating that compact models can capture the majority of the influences affecting structural strain at the wing root.

6.2 Model deployment on independent flights

To evaluate generalisation performance, the models trained on F062 flight data were deployed on the remaining flights listed in Table 1. Performance was assessed using RMSE, fitted line plot (FLP) diagnostics (slope and R^2) and cumulative fatigue damage ratios derived using a power rule method [6].

[4] The entire block of 32,767 models were generated, however only the results for model sizes 2, 4, 6 and 8 are presented here for clarity as larger model sizes did not result in significantly improved accuracy.

A subset of the results from model deployment are given in Table 5. These results show that models derived from the exhaustive parameter selection approach generally exhibited superior deployment performance, with lower errors and more consistent regression behaviour (as indicated by FLP metrics) across flights. SHAP-based models performed well when deployed within regions of the operational envelope similar to their training data but degraded more rapidly under extrapolation in the non-training flights.

Table 5 – F062 training ANN models deployed on other flights

nVars	Flight	Best model				Shapley model			
		%RMSE	Damage ratio	FLP slope	FLP R^2	%RMSE	Damage ratio	FLP slope	FLP R^2
2	F023	5.55	0.80	0.7599	0.8204	8.48	0.00	0.1352	0.0734
	F024	2.20	0.58	0.8040	0.8915	6.08	0.00	0.0834	0.1032
	F062	1.39	0.96	0.9638	0.9642	6.59	0.00	0.1965	0.1961
	F098	12.03	0.34	0.7577	0.9749	12.56	0.00	0.0837	0.0284
	F107	1.81	0.27	0.8013	0.8907	5.05	0.00	0.0547	0.0959
4	F023	5.15	0.97	0.7908	0.8536	4.96	0.97	0.7739	0.8234
	F024	2.20	0.62	0.8060	0.8851	2.51	0.72	0.7640	0.8550
	F062	1.11	1.05	0.9772	0.9773	1.19	0.99	0.9736	0.9738
	F098	10.58	0.41	0.7986	0.9761	9.89	0.19	0.6368	0.9544
	F107	1.93	0.49	0.7860	0.8748	2.10	0.34	0.7633	0.8511
6	F023	9.04	1.30	0.8291	0.5961	15.81	1.18	0.5343	0.0926
	F024	7.83	0.57	0.5587	0.2890	15.73	1.73	0.4122	0.0453
	F062	0.90	1.01	0.9848	0.9849	1.05	0.92	0.9796	0.9795
	F098	8.60	0.36	0.7087	0.8475	7.03	0.36	0.7670	0.8740
	F107	2.32	0.82	0.8353	0.8084	3.10	0.70	0.7881	0.6728
8	F023	3.74	0.77	0.8825	0.7672	11.17	0.65	0.6709	0.3440
	F024	6.87	0.65	0.8906	0.8285	12.28	0.72	0.4552	0.0843
	F062	0.80	0.94	0.9881	0.9881	0.93	0.93	0.9839	0.9839
	F098	7.58	0.17	0.6122	0.8912	17.60	0.45	0.6986	0.4528
	F107	2.72	0.54	0.7817	0.7389	3.26	0.63	0.7752	0.7401

The damage ratio numbers for all 2 parameter Shapley results in Table 5 are all less than 0.005. The poor numerics for this model deployment on F062 are a result of the limited number of input parameters and the exclusion of the NZ parameter.

Visually, an example of the performance metrics for flight F023 wing root strain prediction may be found in Figure 3. The top two plots show the flight trace for both the measured and predicted wing root strain (WPR03AMC1) over time (left) and the FLP of the two strains against one another (right) using the ‘best’ selected 4 parameter input set. In the flight trace plot the measured strain is given in blue while the predicted strain is overlaid in orange. The bottom two plots in Figure 3 show the same metrics but for the Shapley selected 4 parameter input set.

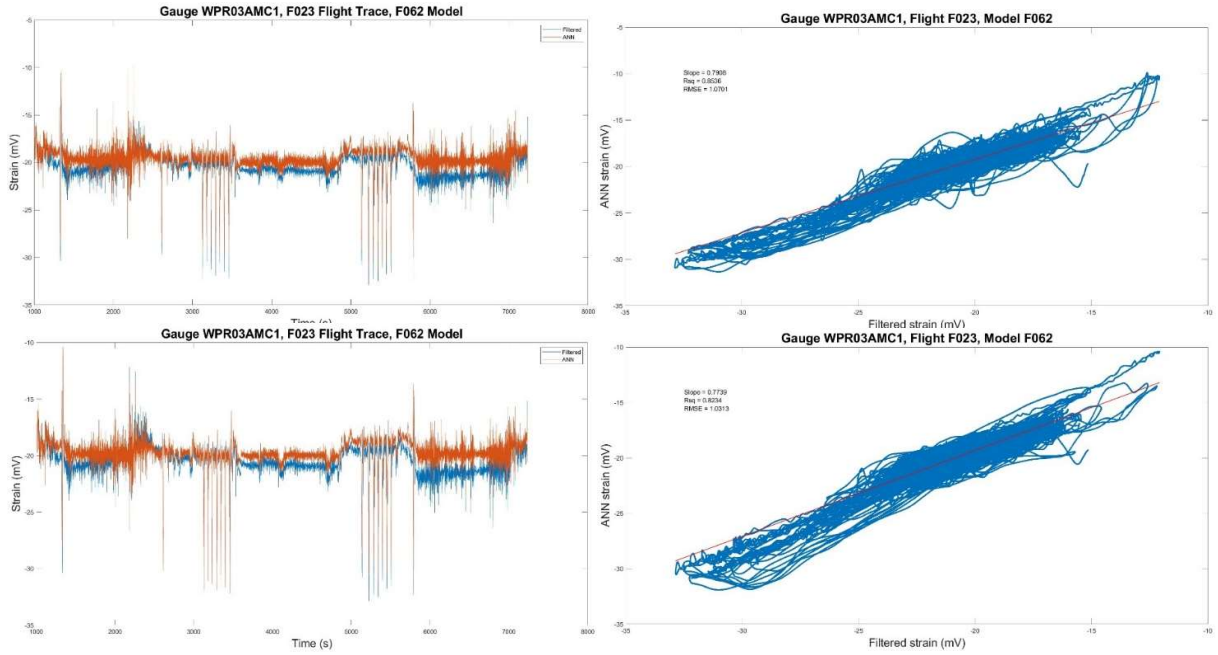


Figure 3 – F023 flight data, 'best' model (top) and Shapley model (bottom), 4 parameter model

Another demonstration of model deployment is shown in Figure 4 using data from flight F098 which had a substantially different payload and fuel distribution as compared to the training flight (F062).[5] In both methods, strain variations are accurately predicted however, a systemic offset is present likely due to differences in aircraft mass distribution, which the model could not account for due to the absence of an explicit cargo and gross weight parameters. This is a useful demonstration of the limitations of any predictive model when taken outside the bound of training data where the model has no ability to accurately extrapolate.

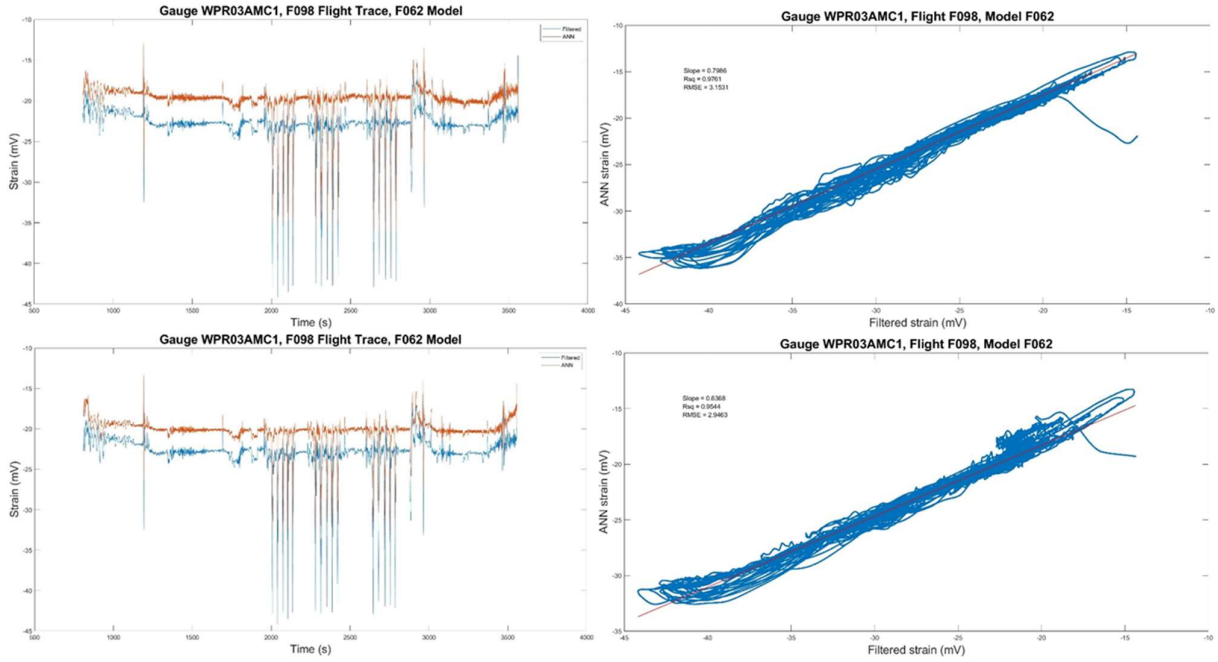


Figure 4 – F098 flight data, 'best' model (top) and Shapley model (bottom), 4 parameter model

[5] Flight 098 has twice the cargo weight of any other flight in the set of five used in this study.

7. Discussion

During model development the exhaustive (best) parameter selection method consistently[6] produced models with lower %RMSE values across all model sizes (Table 4)[7]. The exhaustive method also selected parameters which, based on domain knowledge, more intuitively linked to the drivers of strain at an aircraft wing root.

The generally superior performance of the exhaustive method over the SHAP-based method during model development was not unexpected as the former uses all possible combinations of input parameters to generate the model with the lowest %RMSE. By design, the exhaustive method should produce the "best possible" model and create a baseline against which other methods can be contrasted in terms of both accuracy but also, computational expense. Whilst the exhaustive method is generally more accurate, the degree of improvement in accuracy is not in proportion to the much larger computational effort used to generate the result. In this study, the identification of the best model from the exhaustive method required 264 hours of processing. The SHAP-based method, while slightly less accurate, was able to be completed in just 1.1 hours of processing (roughly two orders of magnitude faster) as it only requires a single ANN to be generated. In contrast, the accuracy differences between the two methods for moderate-sized models are much smaller than two orders of magnitude.

Regarding generalisation, when deployed on independent flight data, the models selected using the exhaustive approach performed better than those selected using the Shapley method (Table 5). Broadly, similar trends to those seen with the %RMSE during model development were also observed across the other performance metrics (damage ratio, FLP slope, FLP R squared) during model deployment. Across both parameter selection methods, and all model sizes, predicting F024 strains shows the least accuracy (overall average FLP slope values furthest from 1) and the widest variation (overall average FLP lowest R^2 values). Damage ratios are generally less than one, indicating that the ANN models result in less damaging fatigue predictions when compared to the fatigue prediction based on the measured strain - the largest exceptions to this is the Shapley 6 parameter model deployed on F024 with damage ratio 1.73. As expected, %RMSE is inversely proportional to FLP slope and R^2 (as the error increases, the slope and R^2 diagnostics decrease and vice versa).

Deployment performance was strongest when test-flight parameter ranges closely matched those observed in the training flight data. Flights with substantial differences in cargo weights and fuel states exhibited increased offsets in the predicted strain. This is likely the reason that the prediction of strain for flight F098 appears to be the most challenging, with RMSE values consistently high across all model sizes and selection methods. These results highlight the importance of training coverage across mass configurations but also the inclusion of parameters, which can inform the model of such differences (e.g. cargo and gross weight). Improvements in training data selection would also be expected to yield improvements in parameter-strain model predictive accuracy across the range of performance metrics when the models are deployed.

8. Conclusion

The results of this study highlight the importance of appropriate selection of input parameters in the development of ANN-based strain prediction models in the development of intermediate-fidelity IATPs. While neural networks are capable of capturing complex relationships, their performance and robustness are strongly influenced by input parameter composition and the training data used.

This study has shown that models developed using the Shapley method of parameter selection are a realistic way to quickly develop parameter-strain relationships within a short timeframe rather than running all possible combinations to discover the absolute best model. Results show that, in development, the exhaustive (best) approach achieved superior prediction accuracy with intuitive input

[6] In three cases the SHAP method did produce slight lower %RMSE values than the exhaustive method.

[7] The differences in %RMSE values between the exhaustive method and the Shapley based approach (for each model size) were very large for the two parameter models, but quickly stabilised as the number of input parameters increased.

parameters, while the Shapley selected models achieved slightly lower performance for moderate model sizes but with a greatly reduced analysis time (2 orders of magnitude quicker). When deployed on independent data, performance metrics (damage ratio, FLP slope and FLP R squared) for the exhaustive approach were superior to the SHAP-based approach however, the improvements were not proportional to the much larger computational expense required under the exhaustive method. These findings demonstrate that SHAP-based parameter selection offers a viable pathway for rapid development of AI-enhanced intermediate fidelity IATPs, while the exhaustive method remains advantageous when maximum accuracy is required. The use of these methods to determine appropriate parameter selection for strain prediction models is easily transferrable to other aircraft platforms and applications; particularly as both methods are model agnostic and thus can be readily applied across a wide range of analyses.

During further work in this field it would be useful to confirm similar outcomes for other airframe structures (empennage, fuselage and nacelle). Improvements to model predictive accuracy through the inclusion of different input parameters (including cross terms [1]) and enhanced selection of training data should also be explored. Finally the use of a fatigue damage based cost function, instead of minimisation of strain prediction error, could also be explored.

9. Acknowledgements

The authors would like to personally acknowledge the contribution of Leonardo SpA in the generation of the test data used in the research presented in this paper. The C-27J FLTP was a highly complex test program which was successfully delivered through close technical collaboration between Leonardo and the Commonwealth of Australia. The efforts of all Leonardo staff who contributed to the success of the FLTP are greatly appreciated.

10. Contact Author Email Address

kate.niessen@defence.gov.au

11. Copyright Statement

The authors confirm that they, and/or their company or organisation, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

References

[1] Reed, S. C. Develoment of a parametrc-based indirect aircraft structural usage monitoring system using artificial neural networks. *The Aeronautical Journal,* Volume 111, Issue 1118, pp 209-230, 2007.

[2] Lundberg, S. M. and Lee, S. I. A unified approach to interpreting model predictions. *31st Conference on Neural Information Processing Systems*, Long Beach, California, USA, 4-9 December 2017.

[3] Cooper, A. Explaining machine learning models: a non-technical guide to interpreting SHAP analysis. Accessed 18 February 2025.

[4] Rusignoli, M. T. *C-27J - I.A.T.P.: General theoretical model.* Alenia Aeronautica, 12/NT/8540/N350C/070015 Issue 1, 2007.

[5] Sagittarius A. C-27J Spartan line drawing, Wikimedia Commons. Accessed 16 April 2026

[6] Anderson, I. A., Franke, D. O., Machin, A. S. and Matricciani, E. *PC9/A Fatigue test validation of the test load sequence Issue B:1070 simulated flight hours.* Defence Science and Technology Organisation, DSTO-TR-0377, 1996.