Naval Information Warfare Center PACIFIC

TECHNICAL REPORT 3183
December 2019

Predicting FA-18 Squadron Readiness and Quarterly Flight Hour Execution Using Machine Learning

Dr. Benjamin Michlin
Dr. Ruey Chang
Rick Cruz
Josh Duclos
Dean Lee
Vincent Siu
Charles Yetman
NIWC Pacific

DISTRIBUTION STATEMENT A: Approved for public release.

Naval Information Warfare Center Pacific (NIWC Pacific)
San Diego, CA 92152-5001

This page intentionally blank.

TECHNICAL REPORT 3183
December 2019

Predicting FA-18 Squadron Readiness and Quarterly Flight Hour Execution Using Machine Learning

Dr. Benjamin Michlin

Dr. Ruey Chang

Rick Cruz

Josh Duclos

Dean Lee

Vincent Siu

Charles Yetman

NIWC Pacific

DISTRIBUTION STATEMENT A: Approved for public release.

Administrative Notes:
This report was approved through the Release of Scientific and Technical Information (RSTI) process in January 2019 and formally published in the Defense Technical Information Center (DTIC) in December 2019.

This reports content represents work performed under Space and Naval Warfare Systems Center Pacific (SSC Pacific). SSC Pacific formally changed its name to Naval Information Warfare Center Pacific (NIWC Pacific) in February 2019.

NIWC Pacific
San Diego, CA 92152-5001

NIWC Pacific
San Diego, California 92152-5001

M. K. Yokoyama, CAPT, USN
Commanding Officer

W. R. Bonwit
Executive Director

ADMINISTRATIVE INFORMATION

The work described in this report was performed by the Command and Control and Enterprise Engineering branch code (64320) of the Business & EIS division code (54320), Naval Information Warfare Center Pacific (NIWC Pacific), San Diego, CA. The Commander, Naval Air Forces (CNAF) Program provided funding for this Basic Applied Research project. Further assistance was provided by Cyber / Science & Technology (Code 71740), Communications & Networks (Code 55180), and Intelligence, Surveillance & Reconnaissance (Code 56000).

Released by
Scott Crellin, Head
Command and Control and Enterprise Engineering

Under authority of
Donna Williamson, Head
Business and EIS

This is a work of the United States Government and therefore is not copyrighted. This work may be copied and disseminated without restriction.

The citation of trade names and names of manufacturers is not to be construed as official government endorsement or approval of commercial products or services referenced in this report.

RJP

EXECUTIVE SUMMARY

Given manning-training-equipment datasets from Naval FA-18 squadrons, a machine learning model for determining the monthly mean number of mission capable jets per squadron is created. This model is then extended and used as an input to create an ensemble of models determining the flight hour execution of a squadron over a three-month period. The ensemble of models is then used to credit squadron performance and readiness, and can correctly classify a squadron's future performance with 75 percent accuracy 90-days in advance.

This page is intentionally blank.

CONTENTS

Figures

Tables

APPENDIX

1. INTRODUCTION

The concept of “readiness” in Naval aviation relies on three key aspects: manning, training, and equipment. “Manning” refers to the available man-power including both the enlisted maintainers and pilots. Manning also includes how experienced the maintainers are. “Training” refers to how well the pilots are trained and the quantity of flight hours executed. Finally, “equipment” refers to the necessary aircraft, supplies, etc. Even without a strict definition, it is immediately clear that impairing any of these three key factors would result in decreased readiness for a squadron. Two key metrics are used in the Naval aviation community to measure the health and readiness of a squadron: mission capable aircraft and flight hour execution. Specifically, a mission capable (MC) aircraft is one which meets or exceeds the minimum requirements to be operated and complete a mission. The MC metric is most useful when viewed in the context of the activity phase, MC entitlement (based on phase and funding), and the size of the squadron (most squadrons are composed of 10 or 12-jets). Similarly, flight hour execution is most relevant when interpreted as a fraction of the total number of flight hours entitled to a squadron. Squadron flight hour execution is evaluated on a quarterly (calendar year) basis and is closely connected to MC.

Since their introduction in 1999 the total number of Naval FA-18 Super Hornets has increased nearly linearly to a current number of almost 600 aircraft[1]. There are two variants of Super Hornets – the single seat E and two-seat F. Additionally, the Navy still uses some of the older FA-18 single-seat C Hornets. The FA-18’s can be in one of several readiness states at any given time: not mission capable for supply (NMCS), not mission capable for maintenance (NMCM), partially mission capable (PMC), or fully mission capable (FMC). Ideally, the number of mission capable aircraft would scale linearly with the number of total aircraft, but this is not the case. Figure 1 shows the number of FA-18’s in inventory and the number of MC aircraft by year. In the late 1990s and early 2000s the relationship between MC aircraft and inventory was as expected, but near 2007 there is a clear deviation and the number of MC aircraft plateaus. This is a well-documented issue and new Super Hornets will continue to be manufactured through at least 2023 [2, 3] while readiness improvements are increasingly becoming the focus of leadership [4]. However, Figure 1 implies that this readiness issue may be more complex. Figures 2a and 2b show an increased number of maintenance man-hours per flight hour required implying that the issue may be due to an aging fleet with aircraft that are more difficult (or time-consuming) to repair or aircraft that need repairs more often. Figure 2 shows increasing maintenance man-hours: showing flight hours per month for the fleet vs year (a), and maintenance man-hours per flight hour vs. year (b).

The ability to accurately monitor and predict readiness is extremely important and non-trivial. Predictive power gives squadrons and decision makers the time and ability to reallocate resources, adjust manning levels, and make smarter decisions before a problem has happened. By creating machine learning models to predict readiness, rather than intuition and human insights, there is also the possibility of discovering unintuitive insights into squadron operations.

The goal of this analysis is to predict FA-18 squadron readiness via monthly MC and quarterly flight hour execution. Further, this analysis provides the ability to monitor readiness as a squadron moves through a quarter. This is accomplished in two steps. First, a model is created to predict the mean number of mission capable aircraft for each squadron on a monthly basis. This model is then extended and used as an input for a second machine learning model which predicts the flight hour execution of a squadron over a quarter. Section 2 reviews the man-train-equip datasets along with their conditioning and transformation, Section 3 delineates the analysis strategy for all machine learning models, Section 4 describes the monthly MC prediction models, Section 5 analyzes and evaluates the quarterly squadron flight hour execution model, Section 6 validates and interprets the machine learning model results, and Section 7 reviews the conclusions and discusses future work.

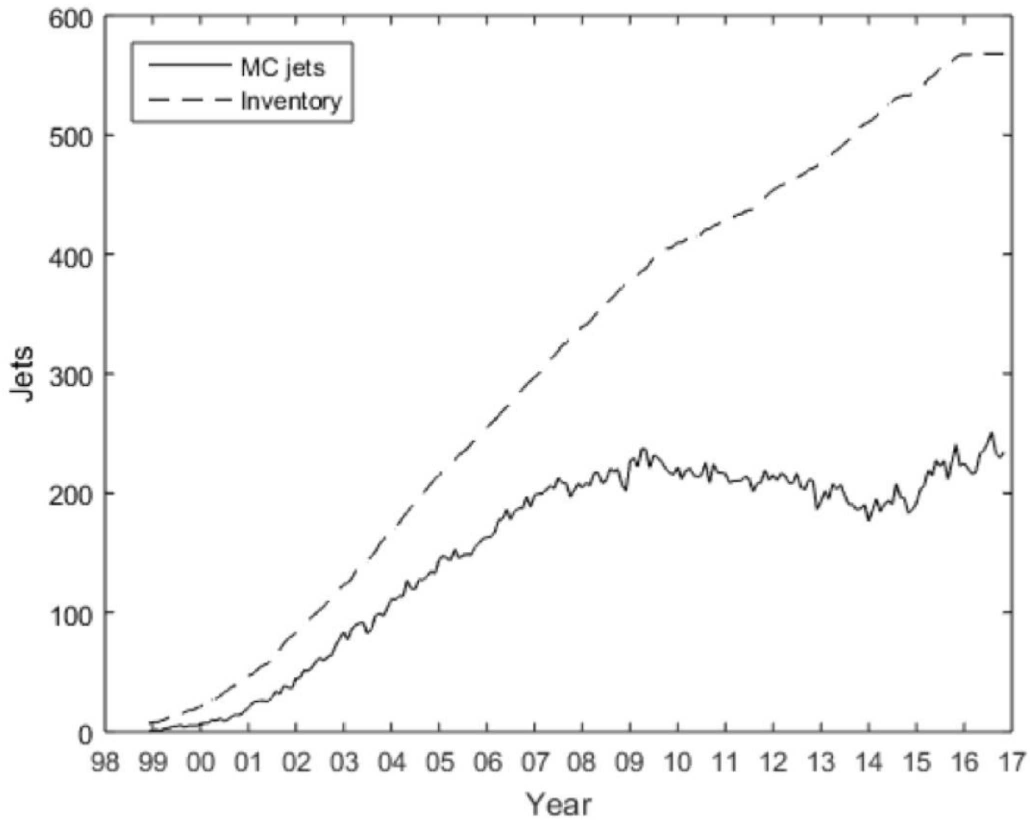


Figure 1. Number of FA-18 jets by year. The number of total jets in inventory increases nearly linearly while the number of MC jets plateaus in $\sim$2007 [5].

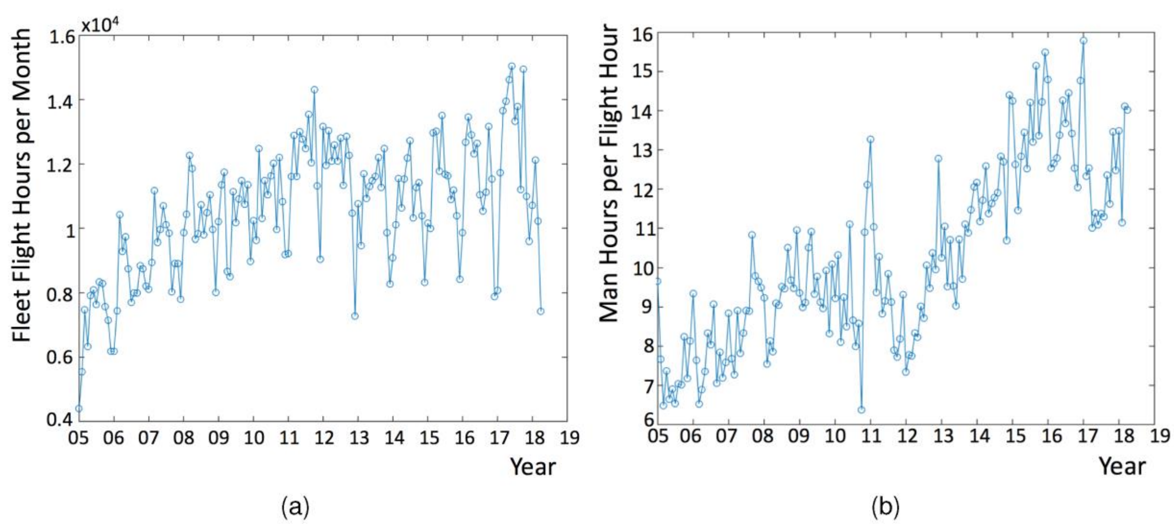


Figure 2. Shows flight hours per month for the fleet vs year (a). Shows maintenance man-hours per flight hour vs. year (b) [5].

2. DATA CONDITIONING AND TRANSFORMATION

Due to the disparate attributes describing the pillars of readiness (man-train-equip), there are many different datasets that are joined to create a master dataset. Table 1 shows the most important data sources used in this analysis. Some of these datasets rely on hand-entered values, others on automatically populated values, and others on subjective attributes. Therefore, each dataset must be handled individually and carefully.

Table 1. Most important data sources and attributes used for analysis.

Man	Equipment	Train
Enlisted Fit/Fill	Deckplate	Flight Hours
Officer Fit/Fill	AMSRR	Training Progression
AMEX 1.0	FHRM	N40 T-Rating
NAE Actualls	NAVAIR Reports	
	Lot	
	NAE Actuals	
	MC/RBA/NMCS/NMCM	

2.1 Data Sources and Aggregation

Since the only features common to all of the datasets are the squadron and date, although the date ranges do not always overlap, these features are used to join the data. There is an added complication in that some datasets record information on a monthly basis while others record daily or weekly. For those that record daily and weekly, aggregate monthly values are created including the mean value, standard deviation, max, min, sum, and count.

2.2 Feature Construction

Once all of the datasets are joined into a master dataset, new features are constructed. Aggregate "entitlements" are constructed features that are meant to be indicative of a squadrons' ideal behavior given size and funding. The pilot entitlement (P_{Ent}) describes the number of pilots that a squadron is entitled to and is defined as

$$P_{Ent} = PAA + 5 \tag{1}$$

where PAA is the primary aircraft authorized or the number of jets in the squadron (usually 10 or 12). The P_{Ent} is then used to create the flight hour entitlement (FH_{Ent}) which is the flight hour execution (FH_{Ex}) goal of a squadron and defined as

$$FH_{Ent} = P_{Ent} \times F_{\%} \times 27 \tag{2}$$

where $F_{\%}$ is the funding percentage as defined by the squadron activity phase. The FH_{Ent} can be calculated on a monthly basis, but squadrons are directed to meet quarterly flight hour execution goals. The flight hour ratio (FH_{ratio}) is then the final readiness metric, and the target to be predicted in Section 3, and is defined as

$$FH_{ratio} = \frac{FH_{Ex}}{FH_{Ent}} \tag{3}$$

where FH_{Ex} are the total flight hours executed within the quarter.

Similar to flight hour entitlements, there are entitlements for the number of ready basic aircraft (RBA)[1] that a squadron is expected to have and is a fraction of their total potential up-aircraft (PAA). The RBA_{Ent} is defined as

$$RBA_{Ent} = \frac{PAA \times F_{\%} \times 0.75}{0.8}. \tag{4}$$

Although they are not directly associated, it is clear that the funding allotment effects the number of mission capable aircraft which directly effects the expected flight hour execution (FH_{Ent}). As with Equation 3, the ratio of mission capable aircraft and the number of entitled MC aircraft is important and defined as

$$MC_{ratio} = \frac{MC}{RBA_{Ent}}. \tag{5}$$

There are many constructed features concerning the manning of the squadrons based on both rate and rank[2]. The key manning features are: basic allowed (BA) which is the number of people that a squadron may request for a given position; onboard or fill ($ONBD$, $Fill$) which is the number of people that a squadron has received for a given position; and fit (Fit) which is determined based on experience working with aircraft and rank. The aggregate constructed features are then the onboard ratio ($ONBD_{ratio}$),

$$ONBD_{ratio} = \frac{ONBD}{BA}, \tag{6}$$

and the fit ratio (Fit_{ratio}),

$$Fit_{ratio} = \frac{Fit}{BA}. \tag{7}$$

These quantities are constructed overall for a squadron and also for individual rates. However, not all rates are assumed to equally influence FA-18 readiness. The rates that correspond directly to aircraft maintainers are collectively known as $DEMOT$ rates. The $DEMOT$ sailor's rates are AD, AE, AM, AO, AT, AME, CS, and PR which correspond to the aviation machinist's mate, aviation electrician's mate, aviation structural mechanic (hydraulics and structures), aviation ordnanceman, aviation electronics technician (safety equipment), aviation structural mechanic (safety equipment), culinary specialist, and aircrew survival equipmentman respectively [6]. The $DEMOT$ onboard ratio ($ONBD_{ratio}^{DEMOT}$) and the $DEMOT$ fit ratio (Fit_{ratio}^{DEMOT}) are constructed exactly as in equations 6 and 7 but restricted to $DEMOT$ rates.

The final set of constructed features attempt to capture the maintenance man-hour information as shown in Figure 2b. Additionally, the following features attempt express the idea of "maintenance capacity" [7]. Maintenance capacity guidelines presume that since MC can be stored over time (neglecting scheduled maintenance) and man-hours can not, quantities communicating information regarding man-hour efficiency and usage may be useful in predicting readiness. To this end, the total number of maintenance man-hours (MMH) per DEMOT sailor ($\frac{MMH}{BA_{DEMOT}}$), the maintenance man-hours per MC hour ($\frac{MMH}{MC_{hour}}$), the flight hours per MC hour ($\frac{FH}{MC_{hour}}$), and the maintenance man-hours per flight hour ($\frac{MMH}{FH_{Ex}}$) are constructed and added to the datasets.

[1] In most cases, RBA and MC are equivalent metrics and the syntax used depends on the context.

[2] "Rate" describes the job function and "rank" describes the level of experience and position in the military hierarchy.

The constructed dataset has 1165 features and 3118 samples (squadron-months). The data was collected from January 2010 to October 2017, includes 35 unique squadrons, and is from 11 unique carrier air wings.

2.3 Data Transformation

As discussed in Section 3, it is essential to construct quarterly datasets in which all data for a single squadron from a calendar-year quarter is contained in a single sample. Also included is historical data from previous quarters for each squadron. To make the datasets as realistic as possible, three datasets are constructed – one for each month in the quarter. Each dataset includes the historical information along with all information that is known on the first day of that month in the quarter. For example, the month two dataset will include all information from the previous month, all information from the previous quarter, but incomplete information for the current month and the next (future) month. Explicitly, a feature like the value of the mean number of MC aircraft for the current month and future months cannot be known, but values from the previous month and previous quarter would be known. However, future information relating to activity phase, funding, onboarding information, age and use of aircraft, etc. is known in advance and may be included. After all transformations are complete, the resulting datasets are shown in Table 2.

Table 2. Final dataset information used for analysis. Here, the number of samples is the number of squadron-quarters there are. Note that the number of features increases as more historical information becomes available for use.

Month in Quarter	Number of Features	Number of Samples
1	4531	728
2	5354	728
3	6178	728

This page is intentionally blank.

3. ANALYSIS STRATEGY

he goal of this analysis is to predict FA-18 readiness as indicated by the quarterly FH_{ratio} and classify squadrons as meeting $\geq 98\%$, $< 82\%$, or in-between of their quarterly FH_{ratio}. Since a squadron's ability to execute flight hours is contingent on their number of MC aircraft, it is also important to predict the mean monthly MC of a squadron on a monthly basis and use this as an input feature for a quarterly FH execution model. This general strategy is shown in Figure 3. Although the features to be used in each model is determined by the model itself via feature importance [8], it is hypothesized by subject matter experts that the manning information, especially for $DEMOT$ sailors, and general aircraft information (e.g., years in service and lot) will be key features in determining a monthly MC prediction. It is also assumed that the monthly MC prediction, historical flight hours, and historical squadron behavior will be good indicators of quarterly FH execution.

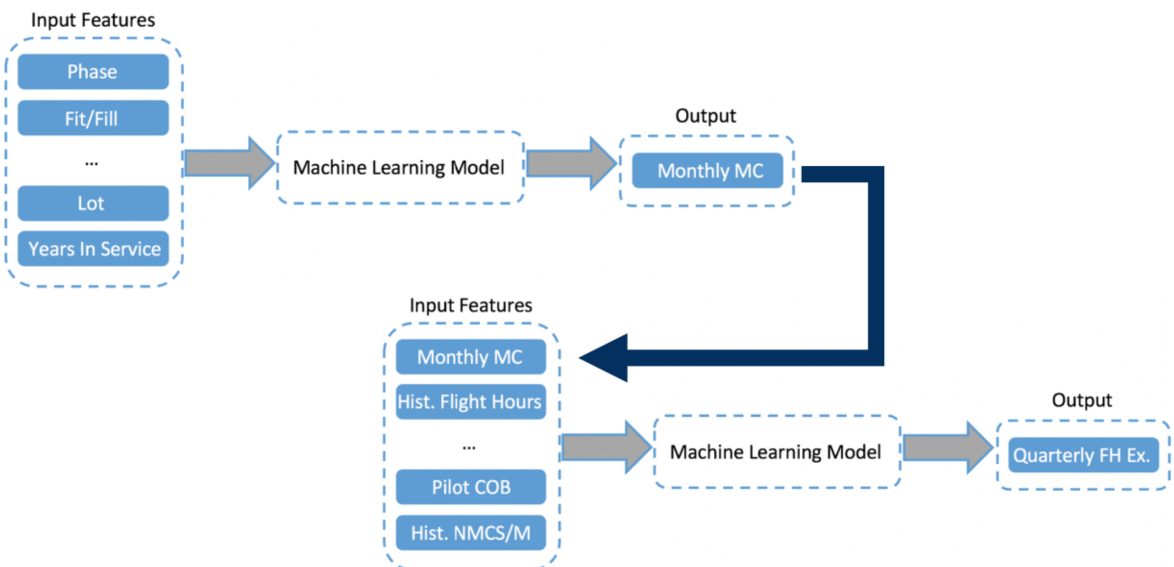


Figure 3. Model strategy. The input features shown are for illustrative purposes only; actual features used are determined by each model via feature importance.

In all machine learning models created, as specified in Sections 4 and 5, the data are randomly split into a training set containing 75% of the data and a hold-out set with 25% of the data. All models are created using the training set and evaluated using the hold-out set.

It is also crucial to be able to track the progress of a squadron as they move through the quarter. To this end, a monthly MC machine learning regression model and a quarterly FH_{ratio} classifier model are created for each month in the quarter. It is necessary to make this distinction between months in a quarter because each month incorporates additional historical information; subsequent months have access to more historical data than the previous month. For example, the MC model and FH execution model for month 1 of the quarter only rely on historical data and current/future data that would be available on the first day of the quarter (see Section 2.3 for details). Then, the MC model and FH execution model for month 2 of the quarter will have all of the same data available to it that month 1 did, but it will have additional information regarding the squadron performance during month 1. Additionally, the month 1 FH model predicts flight hour execution ~90 days in the future while the month 2 model predicts ~ 60 days in the future and the month 3 model predicts ~ 30 days in the future[3]. To this end, a total of 6 models are created and combined to create an ensemble of models.

Furthermore, it is beneficial to have distinct MC and FH models combined to create an ensemble rather than a single combined model. Having multiple models allows for the MC model to be used independently, if desired, and have higher granularity. In many cases it is advantageous to predict a monthly MC. Furthermore, flight hour execution is a quarterly goal for a squadron and, therefore, it may be deceiving to view or predict it on a monthly basis. This is because planned exercises, holidays, weather, etc. are accounted for by individual squadrons. For example, squadrons often over-execute in November to prepare for lower flight hours in December. Lastly, separating the MC and FH models allows for increased interpretability. Thus, individual models for MC and FH predicting individually and on different time-scales is a highly attractive characteristic.

[3] The period that the models predict over changes slightly depending on the length of the months in the quarter that are being predicted.

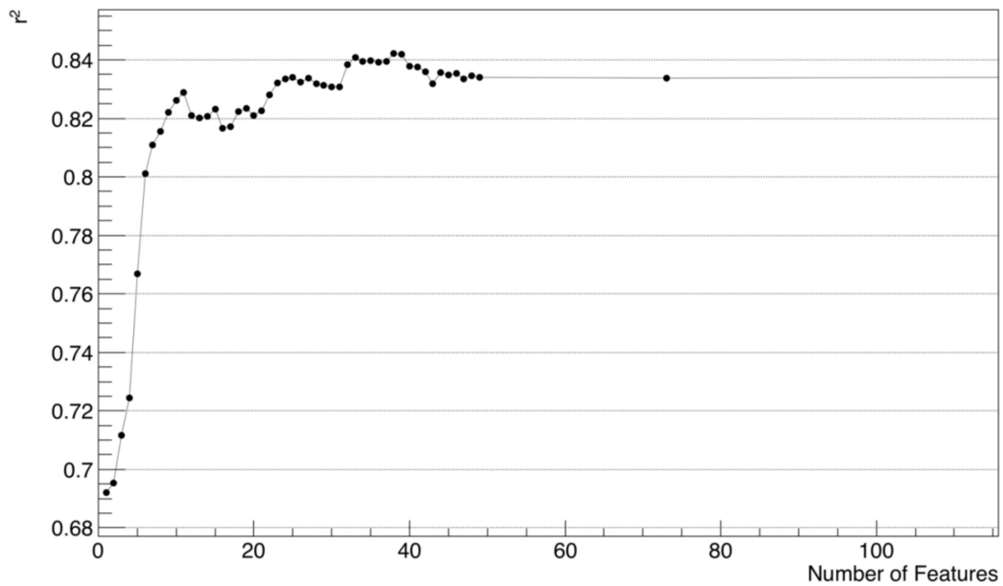
4. MC MODEL

The goal of the MC model, as outlined in Section 3, is to use only the data available on the first day of the month and predict the mean number of MC aircraft for that month. Three MC models are created – one for each month in the quarter. All MC models are created using a training set consisting of a random selection of 75% of the data and are evaluated using a hold-out set consisting of the remaining 25% of the data. All model evaluation metrics are derived from the hold-out set and should be representative of how the MC model will perform and generalize to new data.

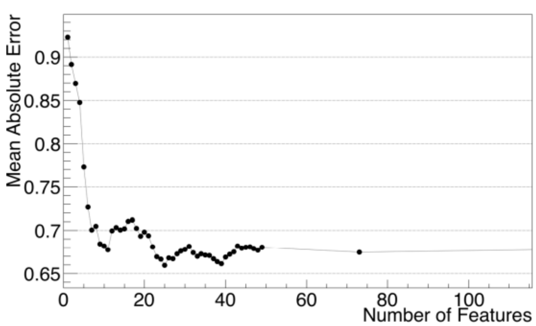
4.1 MC Model Creation and Parameters

A quantile regression forest [9] is used for the creation of the MC model. Quantile regression forests are created such that they have exactly 1 sample in each leaf for every tree in the forest. This means that when a sample is classified as belonging to a leaf, instead of it taking the mean value of training set values that were used when creating the tree it takes the mean value of an actual, pure, sample. In doing so, the prediction of the forest can then be interpreted by looking at the distribution of responses from each tree. The distribution of responses may then be used to determine a prediction interval [10], and the overall response is still taken to be the mean of the tree outputs. A prediction interval is incredibly useful because it provides a way to estimate the likelihood of future observations falling into the specified interval. The distributions of responses may also be interpreted to determine the agreement between trees in the forest. For example, if all of the trees in the forest predict a similar response value, then that prediction should have a higher confidence than a prediction in which trees predict very different values — indicating that the forest is conflicted.

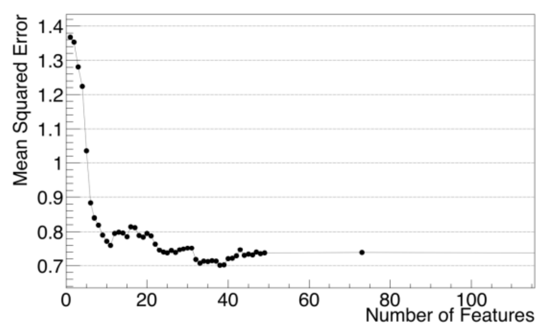
To determine the optimal monthly MC model parameters an ablation study was performed. A model is first created using 500 trees and all available features. A feature importance algorithm [8] is then run and the least important feature is removed. This process is repeated until there are no remaining features, and the model metrics, r^2, mean squared error (MSE), and the mean absolute error (MAE), are recorded for each iteration of the model. The model metrics, shown in Fig 4, are then evaluated to determine the optimal number of features to be used in the model. It is clear that all model metrics plateau after $\sim$40 features are considered. To be well within the plateau region, 200 features are selected for model creation and then hand-tuned to mitigate redundancies and duplications. For example, it would be undesirable for a model to use a feature for FA-18 E's, F's, and then a combined E/F value. This is because redundancies may place undue importance on a feature and introduce biases. It is important to note that only metrics for the month 1 model are shown, but the same procedure is followed for all models. Furthermore, although it would be possible to use every available feature, it is preferred to have a more simplistic model that may be human-interpretable. An interpretable model offers insights into how to influence a squadron's performance and, as outlined is Section 7, may be used to offer an optimized course of action (COA) in the future.



(a)



(b)



(c)

Figure 4. MC regression model metrics. Metrics are shown vs number of features used to create model where (a) shows r^2 values, (b) shows the mean absolute error, and (c) shows the mean squared error. All metrics plateau after $\sim$40 features. Note that metrics are only shown for the month 1 model, but the same procedure is followed for all models. Simple lines are shown to guide the eye.

Once the model features have been curated, the number of trees to use in the forest is examined. Figure 5 shows the r^2 value versus the number of trees in the forest for the month 1 MC model. A plateau is seen after $\sim$200 trees, and 500 trees is selected to be well within the plateau region. No additional alterations are made to the MC model quantile regression forest from the default hyperparameters specified in the scikit-learn package [11].

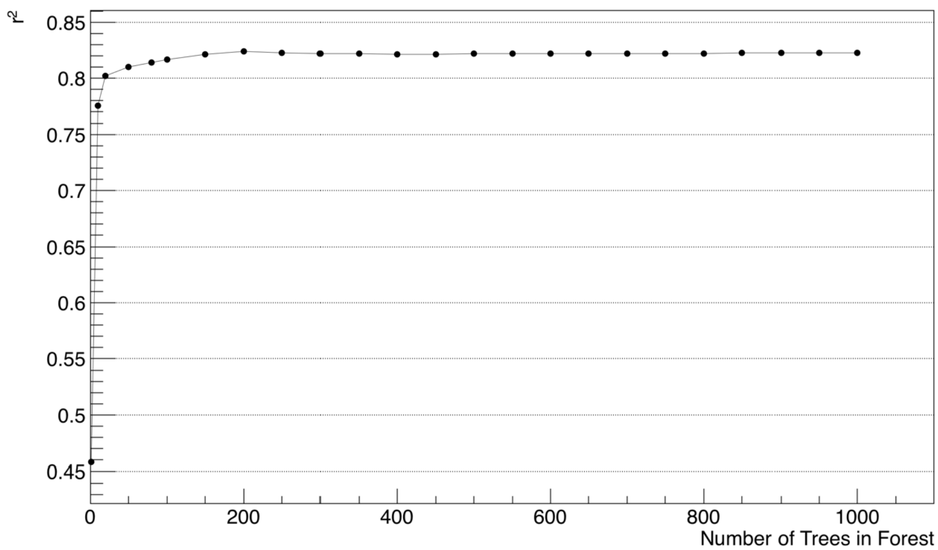


Figure 5. r^2 versus number of trees in forest for month 1 MC model.

4.2 MC Results and Evaluation

The final month 1 MC model features and feature importance is shown in Figure 6. The suffix number signifies the month in the quarter and "pqm" indicates the previous quarter. For example, the feature $mcRatio\_3\_pqm$ is the MC_{ratio} as defined by equation 5 for the third month of the previous quarter. Note that month 2 and 3 models use the same features but include additional historical data from the previous months.

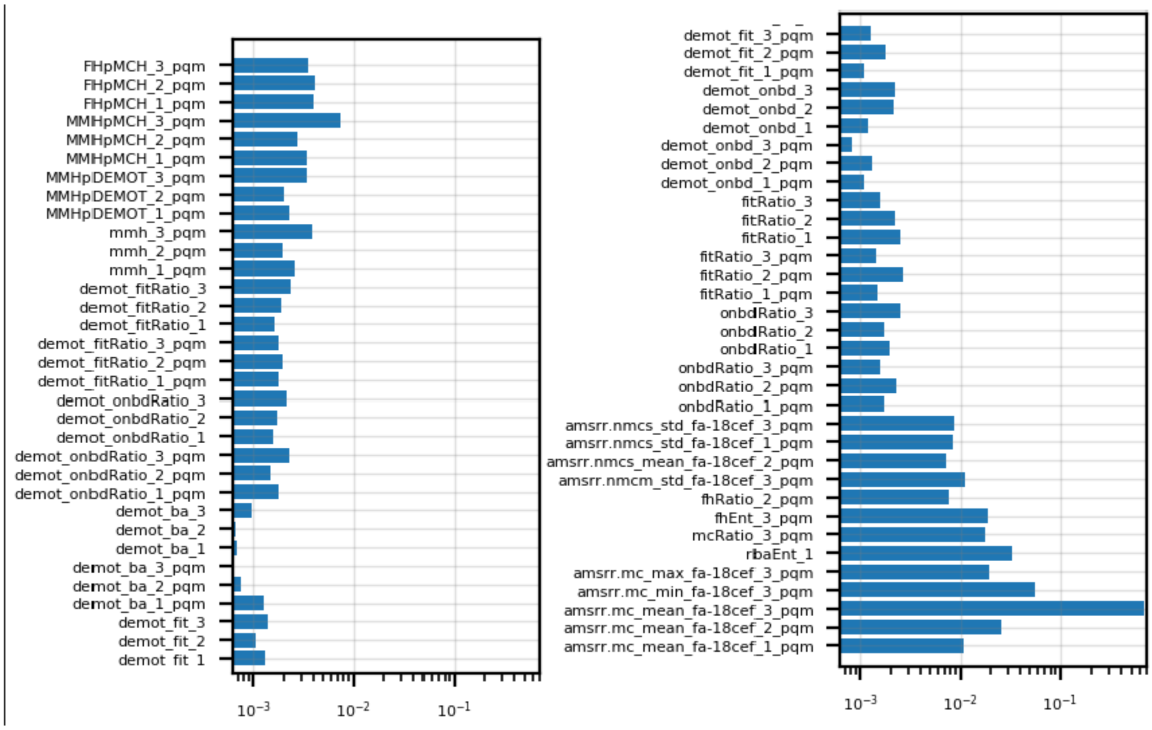


Figure 6. Normalized feature importance for all features used in the month 1 MC model. The suffix number signifies the month in the quarter and "pqm" indicates the previous quarter. The importance, x-axis, is unitless and is log-scale.

The following list details Figure 6 content definitions from the top listed item down starting at the top left:

- $FHpMCH$ is flight hours per MC hour,
- $MMHpMCH$ is maintenance man-hours per MC hour,
- $MMHpDEMOT$ is maintenance man-hours per $DEMOT$ basic allowed,
- mmh is total maintenance man-hours,
- $demot\_fitRatio$ is $\frac{Fit^{DEMOT}}{BA_{DEMOT}}$
- $demot\_onbdRatio$ is $ONBD^{DEMOT}_{ratio}$,
- $demot\_ba$ is BA_{DEMOT},
- $demot\_fit$ is Fit^{DEMOT},
- $demot\_onbd$ is $ONBD^{DEMOT}$,
- $fitRatio$ is Fit_{ratio},
- $onbdRatio$ is $ONBD_{ratio}$,

- $amsrr.nmcs\_std\_fa - 18CEF$ is the standard deviation of the number of not mission capable for supply FA-18 C/E/F aircraft from the AMSRR data,
- $amsrr.nmcs\_mean\_fa - 18cef$ is the mean number of not mission capable for supply FA-18 C/E/F aircraft from the AMSRR data,
- $amsrr.nmcm\_std\_fa - 18CEF$ is the standard deviation of the number of not mission capable for maintenance FA-18 C/E/F aircraft from the AMSRR data,
- $fhRatio$ is FH_{ratio},
- $fhEnt$ is FH_{Ent},
- $mcRatio$ is MC_{ratio},
- $rbaEnt$ is RBA_{Ent},
- $amsrr.mc\_max\_fa - 18cef$ is the maximum value for the number of MC FA-18 C/E/F aircraft from the AMSRR data,
- $amsrr.mc\_min\_fa - 18cef$ is the minimum value for the number of MC FA-18 C/E/F aircraft from the AMSRR data,
- $amsrr.mc\_mean\_fa - 18cef$ is the mean value for the number of MC FA-18 C/E/F aircraft from the AMSRR data.

The most important features are the historical mean MC information, with entitlement information (likely as it relates to funding and phase), fit/fill information overall and for $DEMOT$ sailors, and maintenance man-hour information are also important. The model also does not improve with additional historical data prior to one quarter and does not require planned future information further than the next months in the current quarter. As is to be expected, all models perform equivalently as shown in Table 3. Table 3 also includes information for a generalized MC model discussed in Section 4.3.

Table 3. MC model metrics for each MC model. Metrics shown are r^2, mean absolute error (MAE), and mean squared error (MSE). All metrics are derived from hold-out set.

MC Model Month	r^2	MAE	MSE
1	0.822	0.704	0.791
2	0.792	0.792	1.020
3	0.801	0.734	0.907
Generalized	0.783	0.746	1.082

Figure 7 shows the true versus predicted MC for the month 1 model. The red line shows a perfect prediction. As expected from Table 3, there is very good agreement between the actual and predicted MC. Further, since the predicted MC is from the hold-out set, this is representative of how the MC model can be expected to perform on new data that the model has never seen before. The information from Figure 7 can be viewed in terms of the residual value (predicted - actual MC) and is shown in Figure 8. The red line in Figure 8 is a Gaussian fit to the data and is performed using the MINUIT package [12] as implemented in the ROOT data analysis framework [13]. The χ^2 goodness-of-fit test [14] shows that the residual is well described by a Gaussian centered on 0 with a sigma consistent with ± 1 MC. This shows that the MC model is not biased to over or under-predict MC, and that the hold-out set is a generalized representation of the training set.

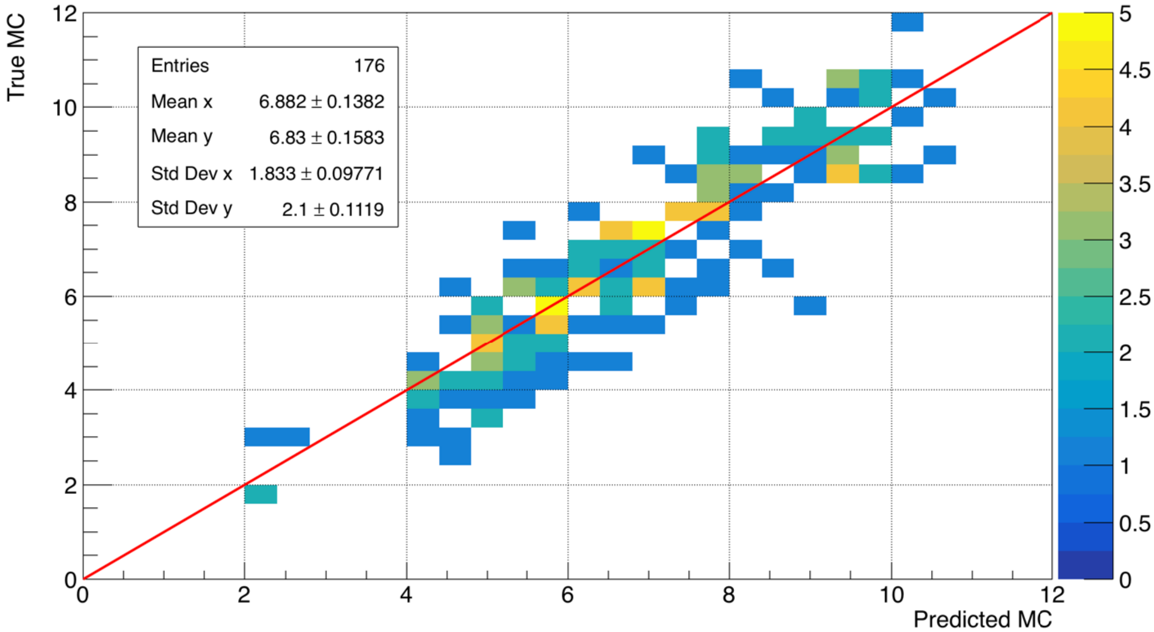


Figure 7. A 2D histogram of true vs. predicted MC for month 1 where the color axis shows the number of entries in each bin. The red line shows a perfect prediction.

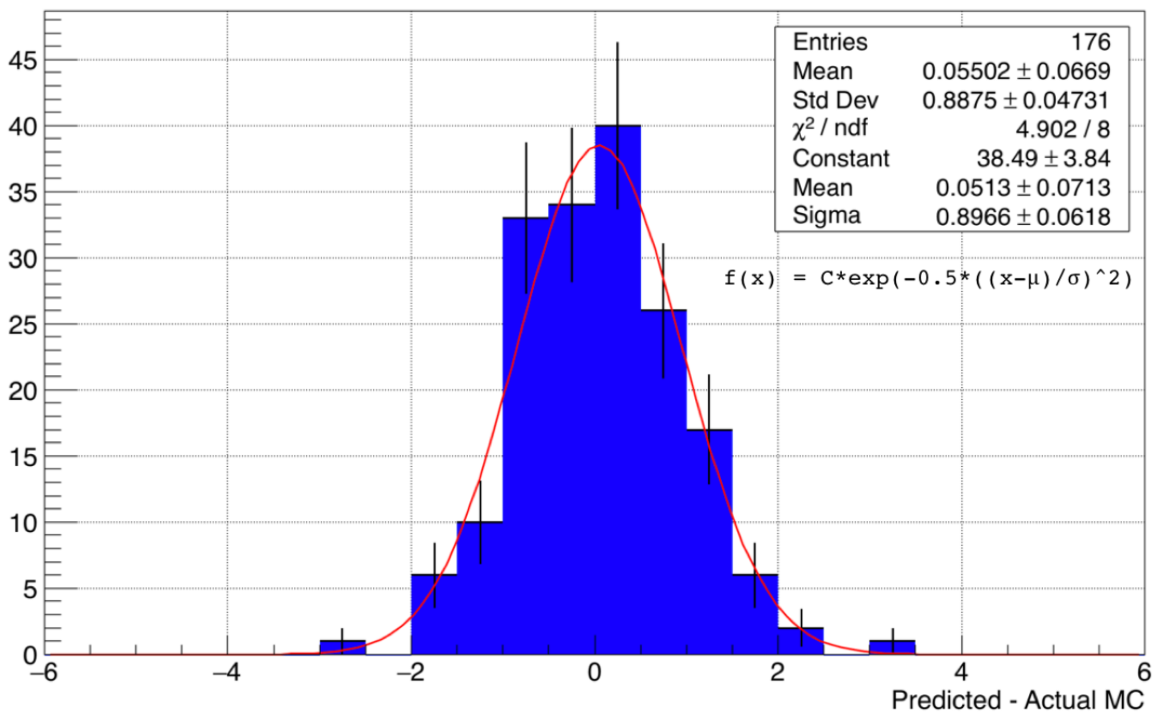


Figure 8. The MC prediction residual (predicted - actual) for month 1. The red line shows a Gaussian fit to the data.

Finally, as discussed in Section 4.1, since a quantile regression forest is used, a prediction interval can be determined for each individual prediction. Figure 9 shows MC versus the sample number for the month 1 MC model using the hold-out set of data. The samples are ordered such that the actual MC value, shown in red, is increasing. The blue points are the predicted MC values connected by a simple line including an extremely small statistical uncertainty from the standard uncertainty on the mean of the predictions from the quantile regression forest. The green and yellow shaded areas show the $\pm 1\sigma$ and $\pm 2\sigma$ prediction intervals respectively. Figure 9 again shows that the predicted and actual MC values are extremely similar. However, the prediction interval is generally smaller, and the predictions more accurate, when the actual MC is mid-range with less predictive power when MC is extremely low, $<\sim 4$, or very large, $>\sim 10$. This is likely due to fewer data samples for the edge cases. Additionally, it is clear that the prediction intervals provide valuable insight into the confidence of the predicted value. In this way, the model itself can provide guidance into how its results should be interpreted.

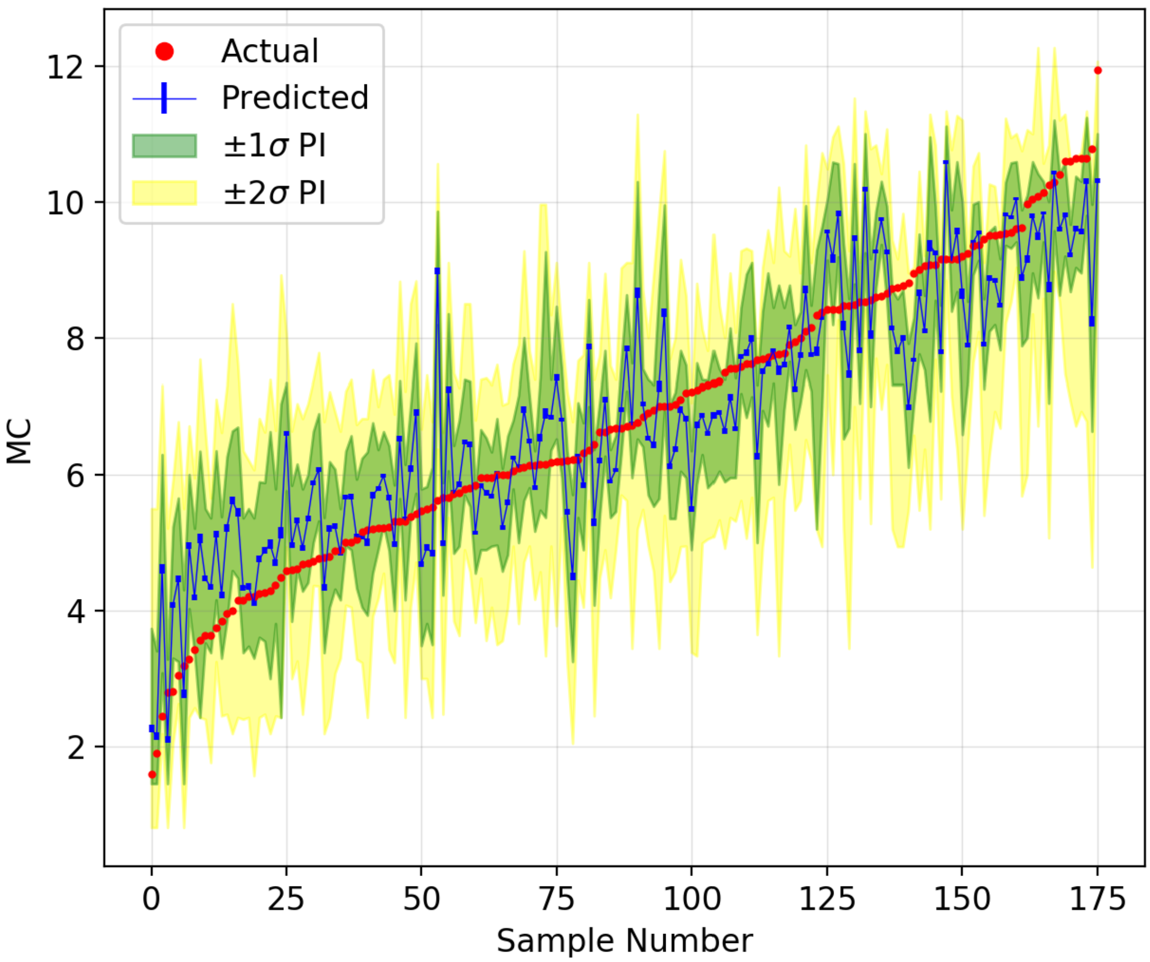


Figure 9. MC vs. the sample number for MC model month 1. Red dots show the actual MC and the sample numbers are ordered for increasing MC. Blue points show the predicted MC including statistical uncertainty. The blue lines are connected with a simple line to guide the eye. Green and yellow shaded areas show $\pm 1\sigma$ and $\pm 2\sigma$ prediction intervals respectively.

4.3 Generalized MC Model

In certain circumstances it may be beneficial to have a single generalized MC model rather than an MC model specific to each month in a quarter. It may also be useful to have an MC model that does not rely on historical or future data. For example, if there is a new squadron or if enough changes have been made to a squadron to make it discontinuous from its historical data then a more general MC model will be extremely valuable.

Following the same procedure as the previous MC models, as outlined in Section 4.1, a generalized MC model has been created. Since there is no future or historical information, the dataset has also changed. There are now only 221 features, but 2914 samples (squadron-months). As with all models, the machine learning models are trained using a random 75% selection of the data and evaluated using a hold-out set consisting of the remaining 25% of the data. The performance of the generalized model is shown in Table 3. Even though the generalized model does not include historical or future information, it performs extremely well and is competitive with month-specific models.

The features used in the generalized MC model and their relative importance is shown in Figure 10. Features used in the generalized model but not the month-specific models include $ONBD$ and fit information for the rates YN, NC, AD, MA, PS, AM, and AN. These rates, some of which are $DEMOT$ sailors, correspond to the yeoman, Navy counselor, aviation machinist, master-at-arms, personnel specialist, aviation structural mechanic, and airman respectively. The Fleet Readiness Training Plan (FRTP) phase is also now included, likely as it is related to funding and entitlements. It is also interesting that when maintenance man-hour information is not available (because this information is only available as historical information) the model chooses to include lot information which is related to when the aircraft was manufactured.

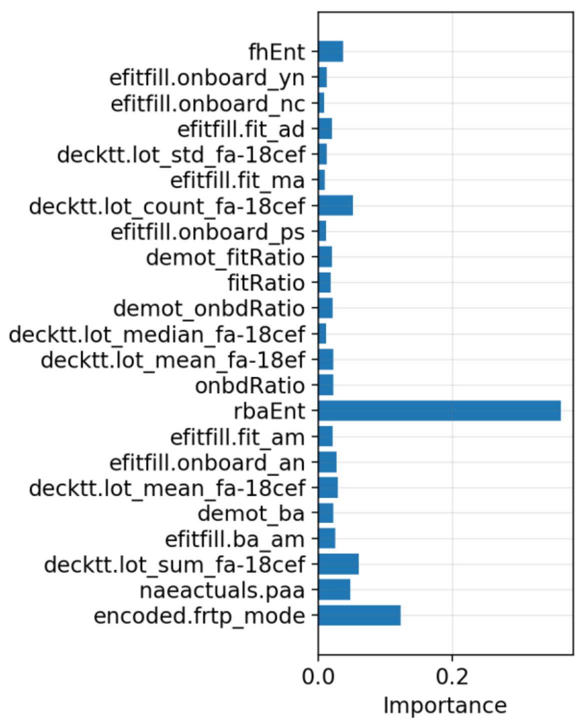


Figure 10. Normalized feature importance for all features used in the generalized MC model.

The MC versus the sample number for the generalized model is shown in Figure 11. Interpretation is the same as in Figure 9. Figure 11 shows that the model performs extremely well and that a generalized model is much more likely to accurately predict very high and low actual MC values. This is likely due to the increased size of the dataset.

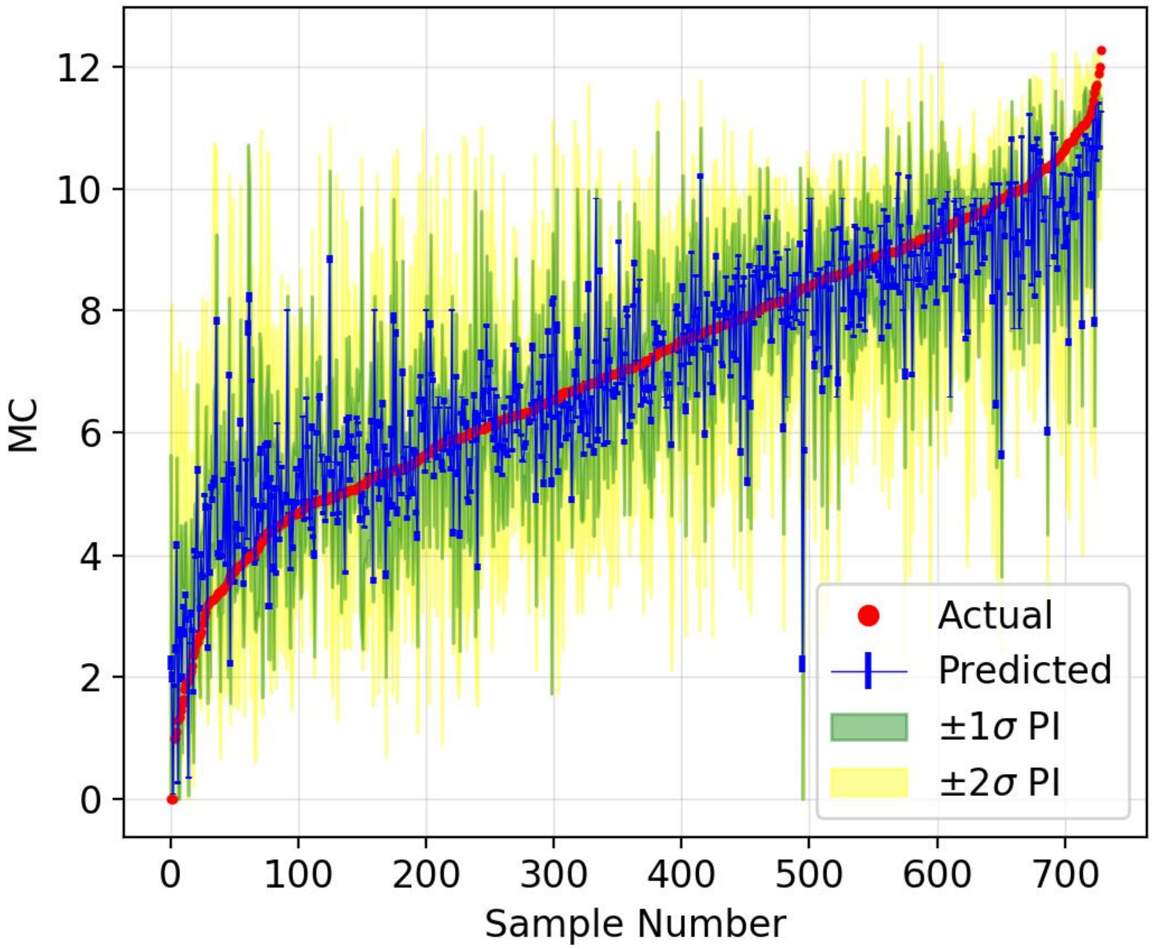


Figure 11. MC vs the sample number for generalized MC model. Red dots show the actual MC and the sample numbers are ordered for increasing MC. Blue points show the predicted MC including statistical uncertainty. The blue lines are connected with a simple line to guide the eye. Green and yellow shaded areas show $\pm 1\sigma$ and $\pm 2\sigma$ prediction intervals respectively.

The generalized and month-specific MC models have shown that each squadron, which was original thought to be unique, actually behaves in a consistent and predictable way. As a standalone result, the MC models are a powerful tool to predict a key aspect of readiness.

This page is intentionally blank.

5. COMBINED MC-FH EXECUTION MODEL

monthly MC models from Section 4 are then used as an input to create an ensemble flight hour execution model called the MC-FH model. The distribution of quarterly flight hour ratios (FH_{Ex}/FH_{Ent}) is shown in Figure 12. Because the flight hours executed is incredibly complicated and relies heavily on undeterminable attributes, like weather, a probabilistic classification model is created rather than a regression model. The goal is then to predict whether a squadron will meet $\geq 98\%$ of their flight hour entitlement (green), less than 82% of their entitlement (red), or in-between (yellow). These classes are shown in Figure 12. As with the MC model, an MC-FH model is created for each month in the quarter so that a prediction for the squadrons' flight hour ratio at the end of the quarter can be made on the first day of each month in the quarter. Each model is trained on a random selection of 75% of the data and evaluated on a hold-out set composed of the remaining 25% of data. Note that this is the exact same training and hold-out set used in the creation of the MC models. Having separate MC-FH month models allows the progression of a squadron to be tracked as it moves through the quarter.

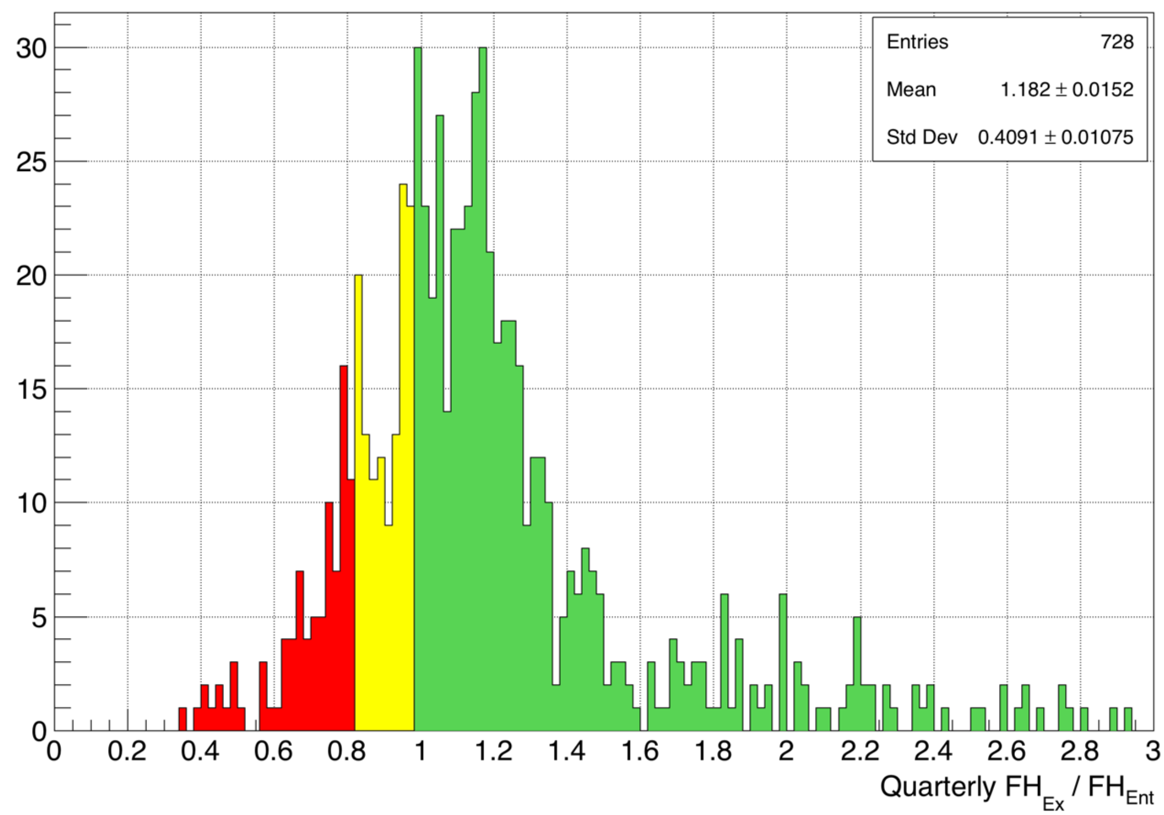


Figure 12. Histogram of the quarterly flight hour ratio is shown. Delinieations for $\geq$ 98% of their flight hour entitlement (green), less than 82% of their entitlement (red), or in-between (yellow) is shown.

5.1 Model Creation

The first step in creating the MC-FH model is to include the predicted MC as an input feature for the MC-FH model to select. Then, a random forest probabilistic classifier with 500 trees that is able to use all features (see Section 2) is trained. As with the MC model, a feature importance algorithm is run and the least important feature is removed from the dataset. This process is repeated until there are no remaining features. Machine learning metrics for each model are recorded and shown in Figure 13. Figure 13a shows the F1 score [15] versus the number of features used in the model, and Figure 13b shows the model accuracy versus the number of features used in the model. Both metrics show volatility when a low number of features is used and then plateau as the number of available features increase. The most important 200 features are selected and then hand-tuned to mitigate redundancies and feature conflicts.

5.2 Model Results and Evaluation

The final feature list for the month 1 MC-FH model shown in Figure 14 has many interesting features – from the top:

- the predicted and historical MC ratio,
- the number of pilots currently on board (COB),
- the historical flight hour information
- the allocation information (related to funding),
- the historical flight hour ratios,
- the mean number of partially mission capable aircraft,
- the mean number of not mission capable for supply aircraft,
- the mean number of not mission capable for maintenance aircraft,
- the predicted and historical mean MC aircraft,
- the number of FA-18 C/E/F's in reporting (IR),
- the number of fully mission capable aircraft,
- the number of aircraft assigned to the squadron.

The mean MC predicted by the MC model and the associated MC ratio are the two most important features. This is reassuring and increases trust in the MC model. The remaining features are entirely related to the state of the squadrons' aircraft, funding, and the number of pilots. The feature list tells an intuitive story: if a squadron has the people, jets, and funding, they will execute their flight hour entitlement. It is also worth noting that historical data earlier than the previous quarter and future quarter data is unnecessary. The same feature list with additional historical information is used for the month 2 and month 3 ensemble MC-FH models.

The MC-FH model results for all three MC-FH month models, as evaluated on the hold-out dataset, are shown in the form of confusion matrices [16] in Figure 15. These results are also summarized using their precision, recall (accuracy), and F1-score [17] in Tables 4, 5, and 6. It is immediately apparent that because the classes (red, yellow, green) are unbalanced there is bias in the model. Specifically, the model is more

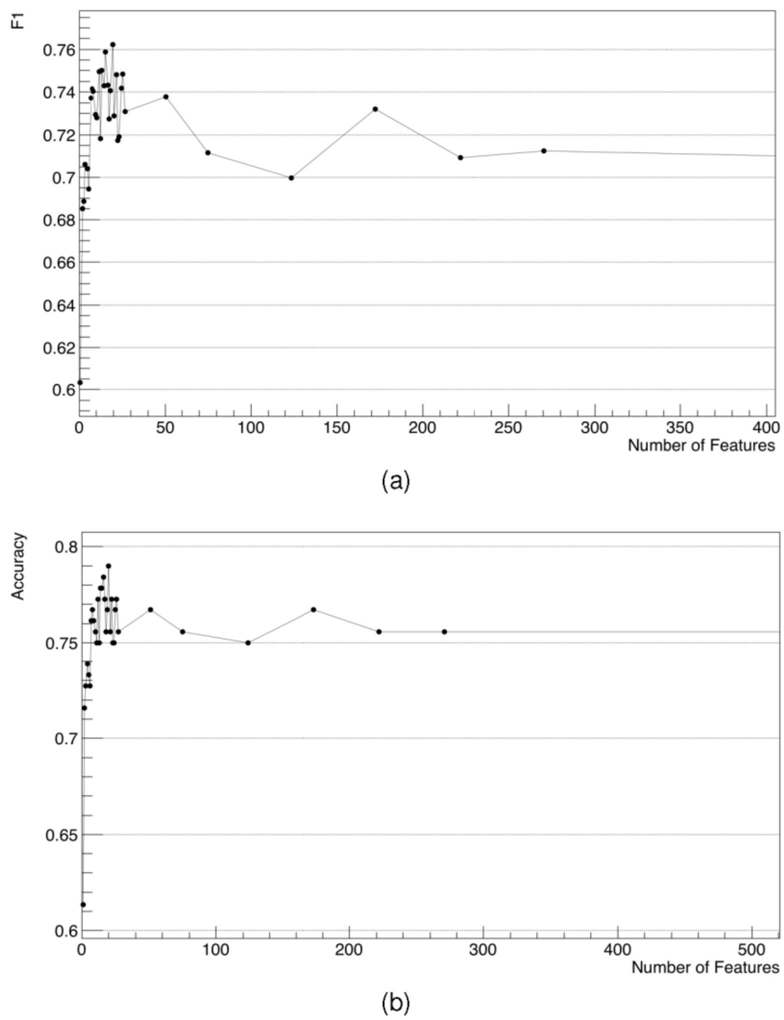


Figure 13. Machine learning metrics shown for month 1 MC-FH model created using variable numbers of features. (a) shows F1 score vs. number of features used in month 1 MC-FH model. (b) shows accuracy vs. the number of features used in month 1 MC-FH model.

likely to predict green than any other class. Also, as expected, the MC-FH models improve as the quarter progresses. Again, this is likely because each successive model is predicting a shorter distance into the future. However, even for the month 1 MC-FH model, predicting the flight hour ratio $\sim$90 days in advance with an average accuracy (recall) of 75% is an extremely powerful result. Further, since this is a multi-class classifier, an accuracy of 75% is astounding.

A multi-class random forest also allows for a probabilistic interpretation of the results. Instead of a single prediction, the MC-FH models may give each class (green, yellow, and red) a probability by comparing the number of trees that predicted each class. These probabilities provide decision makers more information and insight prior to a decision being made. The probabilistic interpretation of the MC-FH model is discussed in Section 6.2.

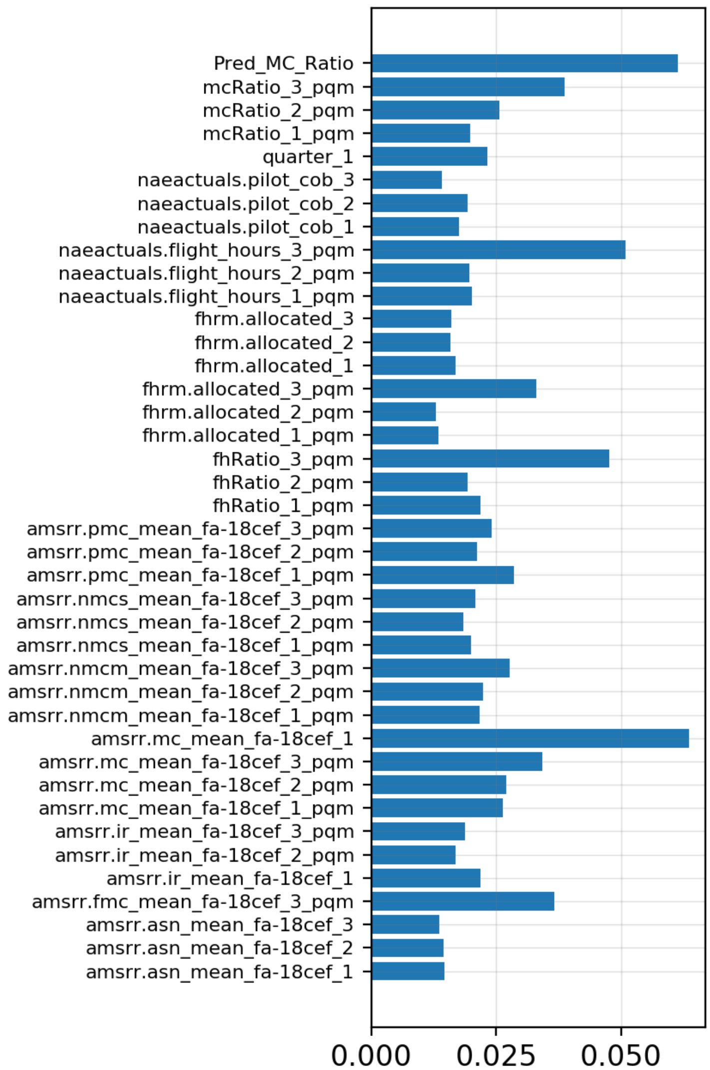


Figure 14. Normalized feature importance for the month 1 MC-FH model.

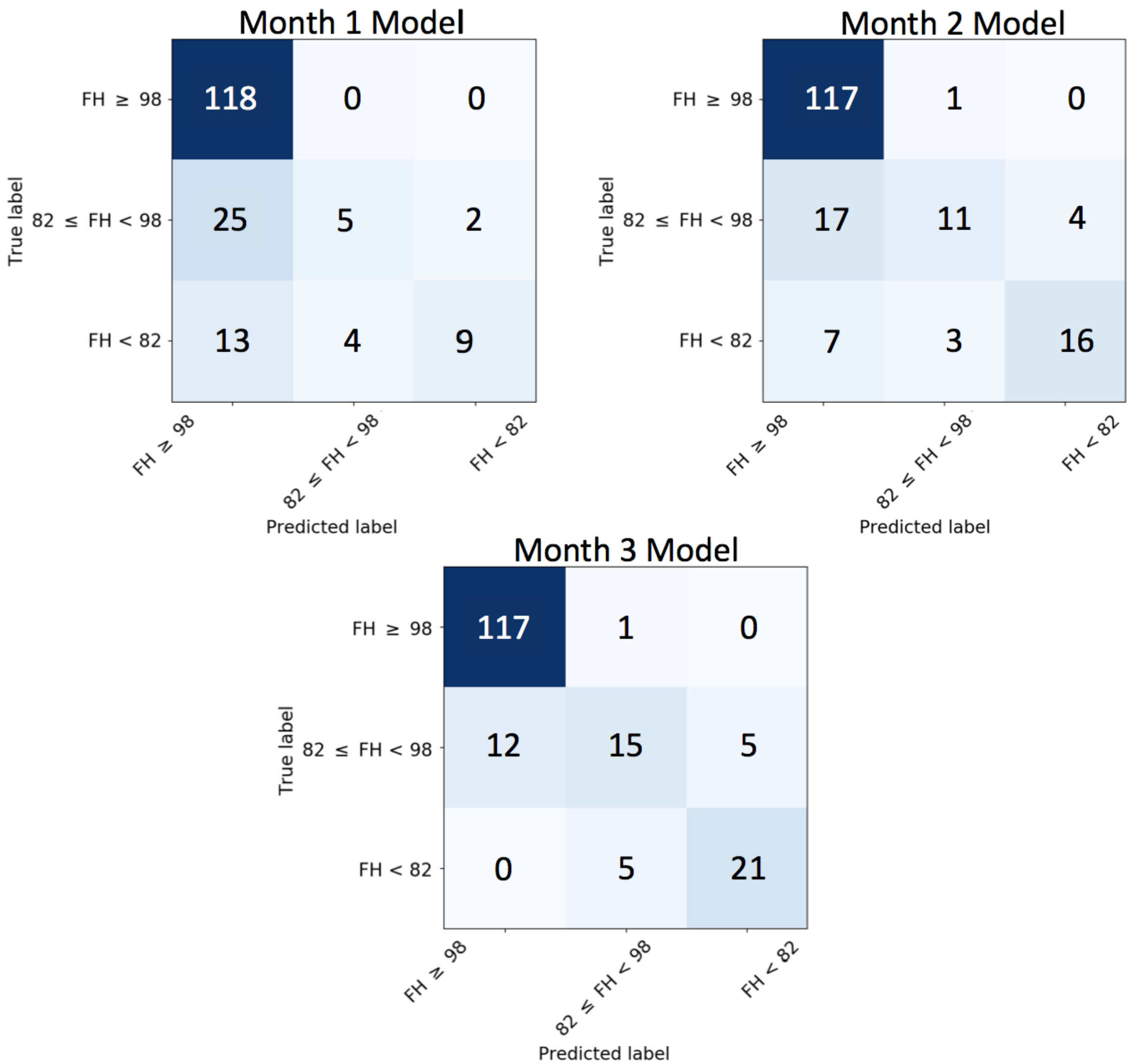


Figure 15. Confusion matrices for all three MC-FH models.

Table 4. Month 1 MC-FH model metrics. Support is the number of events from each class in the hold-out set.

Class	Support	Precision	Recall (Acuracy)	F1
Green	118	0.76	1.00	0.86
Yellow	32	0.56	0.16	0.24
Red	26	0.82	0.35	0.49
Average Total	176	0.73	0.75	0.69

Table 5. Month 2 MC-FH model metrics. Support is the number of events from each class in the hold-out set.

Class	Support	Precision	Recall (Acuracy)	F1
Green	118	0.83	0.99	0.90
Yellow	32	0.73	0.34	0.47
Red	26	0.80	0.62	0.70
Average Total	176	0.81	0.82	0.79

Table 6. Month 3 MC-FH model metrics. Support is the number of events from each class in the hold-out set.

Class	Support	Precision	Recall (Acuracy)	F1
Green	118	0.91	0.99	0.95
Yellow	32	0.71	0.47	0.57
Red	26	0.81	0.81	0.81
Average Total	176	0.86	0.87	0.86

6. VALIDATION AND INTERPRETATION

With all three MC-FH ensemble models created, the models should now be validated and interpreted. Specifically, any biases in the model should be identified and the resulting predictions should be interpreted with care.

6.1 MC-FH Model Validation

To ensure there are no obvious biases introduced due to model selection or choice of hold-out and training sets, the model behavior for different subsets of the data must be studied.

First, Figure 16 shows the accuracy of the MC-FH model on the hold-out set for each month in each quarter. All months have similar behavior wherein the accuracy of the MC-FH model improves as the quarter progresses. Further, each model month in each quarter has similar accuracy. However, there are several minor differences. The accuracy of the first quarter is slightly lower than that of the other quarters. This is expected, understood, behavior and is due to conditions that influence flight hour execution and readiness but are not accounted for in the data. For example, the weather in winter (quarter 1) has a large influence on whether or not a squadron can execute their flight hours. Holidays, lengths of months, scheduled operations, etc. are also not included in the model creation but cause slight discrepancies between the model behavior over the different quarters.

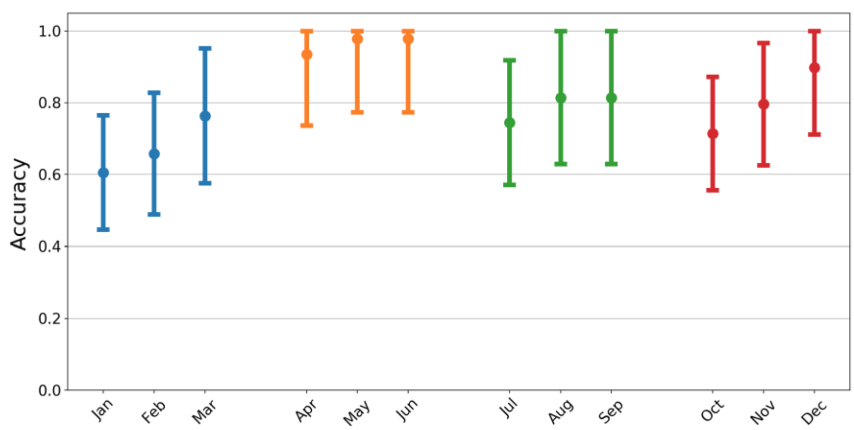


Figure 16. Accuracy vs. month for every month shown including statistical uncertainty.

Second, the model performance is measured for each individual squadron. This is meant to check that the MC-FH model accurately predicts the readiness of all squadrons equally well. The same analysis is performed as in Figure 16; however, since there are 35 squadrons and the hold-out set is rather small, statistical uncertainty dominates and no concrete statements may be made. Although, there does not seem to be a bias present relating to the individual squadrons, carrier air wings, or any similar categorization. This is because if those features were predictive then they would not have been discarded during the feature selection process. Furthermore, the model was shown to generalize well as discussed in Section 5.2.

6.2 Probabilistic Interpretation

As previously described, the MC-FH model output is a vector of probabilities corresponding to each class (red-yellow-green). The highest class probability corresponds to the most likely class as predicted by the model, and is called the prediction confidence. Ideally, the prediction confidence would correspond directly the probability of a correct prediction. However, this is not the case because of systematic uncertainties associated with shared features between trees in the random forest (tree dependence), Monte Carlo noise due to a finite number of trees, inaccurate training data from the MC model, and other uncontrollable biases [18, 19]. These biases and systematic uncertainties are encapsulated in Figures 17 and 18. Figure 17 shows the probability of a correct prediction versus the prediction confidence including statistical uncertainties for the month 2 MC-FH model. The probability of a correct prediction and the prediction confidence are extremely similar and nearly linear which means that the systematic uncertainties must be small. Figure 18 shows the probability of a correct prediction versus the prediction delta including statistical uncertainty. Here, the prediction delta is the difference between the most likely class probability (the prediction confidence) and the second most likely class probability. The prediction delta encapsulates all of the information contained in the probabilistic result and can give an impression of how distinctly the model classifies a sample. For example, if the model predicts green-yellow-red as 50-25-25 that is a much different prediction than 50-49-1. In both cases the prediction confidence is 50%, but in the first scenario the prediction delta is 25 and in the second case the prediction delta is 1.

Additionally, imperfections from the MC model prediction are propagated through and will influence the final MC-FH prediction. To account for this uncertainty, when a prediction is made the MC value is varied within $\pm 1\sigma$ with a step-size of 0.1. The standard uncertainty (the standard deviation divided by the square-root of the number of points) is then applied to the nominal value. This uncertainty has very little effect on the final MC-FH prediction, likely because the complete MC-FH model has compensated for the uncertainty in the MC prediction during training.

Overall, when the prediction from the MC-FH model is being interpreted Figures 17 and 18 allow for better, smarter decisions to be made. In practice, decision makers may set their own confidence thresholds for action and different thresholds may be set under any number of conditions. In this context, a probabilistic interpretation of the MC-FH model allow the results as shown in Section 5.2 to create an even stronger final result.

6.3 Statistical Interpretation

Although most of the features used are only statistically meaningful when combined in complicated ways, MC is such a strong indicator of readiness that it has a direct correlation to flight hour execution. Figure 19 shows the probability of a class (green-yellow-red) versus the MC ratio for each class. Figure 19 uses all available data and the MC from the final month in the quarter, though other months show similar behavior. It is clear that the probability of a squadron executing $\geq 98\%$ of their flight hour entitlement falls rapidly if their MC is below their entitlement. Further, if their MC ratio falls below 0.8 then it becomes extremely unlikely that a squadron will meet their flight hour entitlement. Surprisingly, if a squadron has more MC aircraft than their MC entitlement, they see diminishing returns in the probability of executing $\geq 98\%$ of their flight hour entitlement.

Other features of the MC and MC-FH model are separated into "levers" and "drivers". Levers are properties that can be changed, for example, increasing funding, transferring maintainers, etc. Drivers are features which influence readiness but cannot be changed. Examples of drivers include the age of the aircraft and the aircraft years in service. It would be desirable to interpret how the model is influenced by changes to less directly correlated features, specifically the levers. To this end, embedded simulation [20] is used to adjust the levers and monitor the model response. No human-interpretable result is found, but the method and findings are discussed in Appendix A.

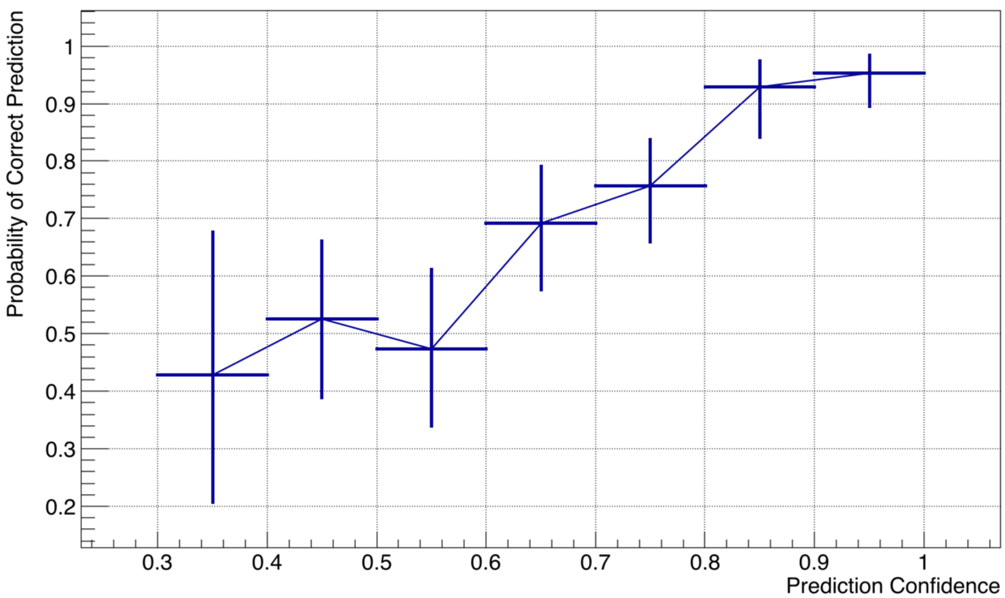


Figure 17. Probability of correct prediction versus the prediction confidence including statistical uncertainties. The prediction confidence is the probability assigned to the most likely class by the MC-FH model. Results are shown for the month 2 MC-FH model, but all models show similar behavior.

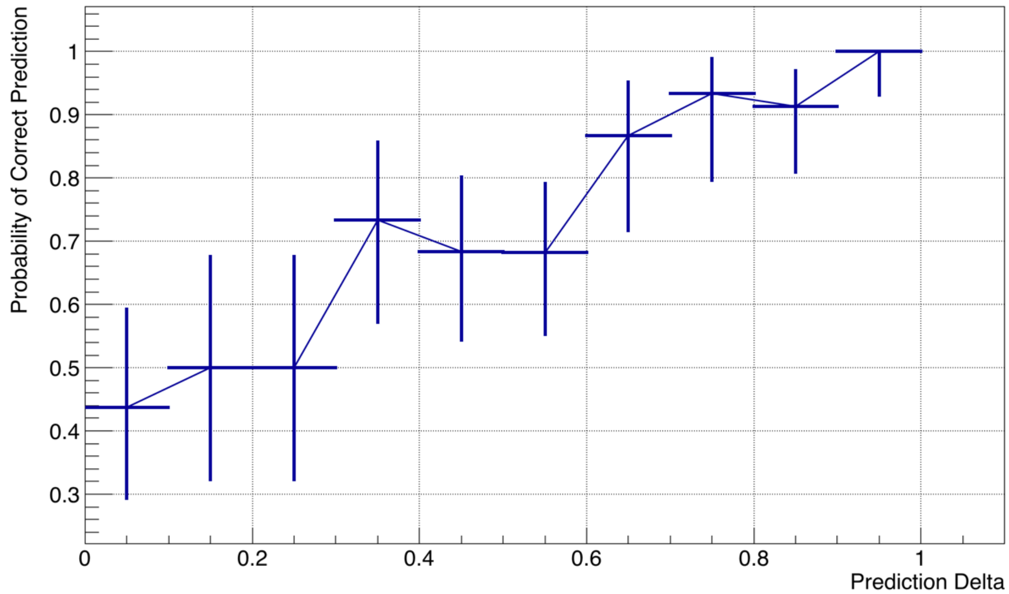


Figure 18. Probability of correct prediction versus prediction delta including statistical uncertainties. The prediction delta is defined as the difference between the most likely class and the second most likely class. Results are shown for the month 2 MC-FH model, but all models show similar behavior.

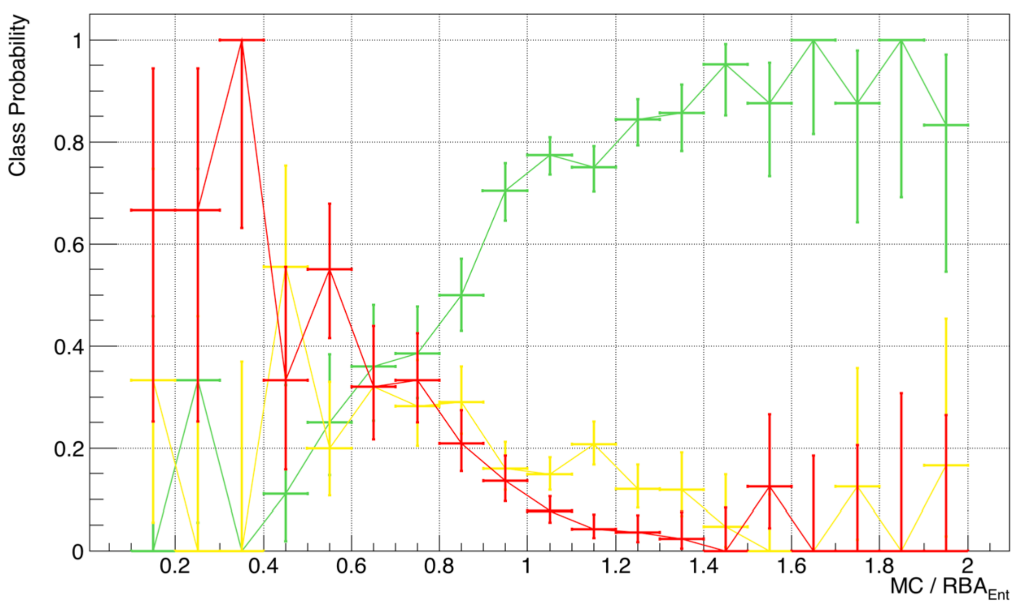


Figure 19. The probability of each class (red, yellow, green) versus the MC ratio including statistical uncertainties. The line colors correspond to their respective classes: $\geq$ 98% of their flight hour entitlement (green), less than 82% of their entitlement (red), or in-between (yellow).

7. CONCLUSIONS AND FUTURE WORK

In conclusion, using man-train-equip datasets, an MC aircraft regression model was created as an input for a squadron quarterly probabilistic flight hour execution predictive model. The ensemble MC-FH model is able to accurately predict the flight hour execution of a squadron as a fraction of their flight hour entitlement with 75% accuracy on the first day of the quarter – 90 days in advance. Separate MC-FH models were created for each month in the quarter in order to incorporate additional historical information, for a total of 6 individual models. By utilizing each model, it is possible to track the readiness and predicted behavior of a squadron as they progress through a quarter. Additionally, a statistical interpretation of the results provides insights into the MC ratio and how a squadron is impacted by fewer or additional MC aircraft.

The generalized MC model may also be considered as a standalone product. Although it does not predict readiness more than 1 month into the future, it may be used as a gauge to measure the current capacity of a squadron. Furthermore, the generalized MC model may be combined with the statistical interpretation of the MC-FH model to quickly assess a squadron's readiness.

Future work includes making certain that the MC-FH models perform well when different hold-out sets are chosen. For example, a hold-out set composed of date range or a never-before-seen squadron. Also, new data and additional data sources may be incorporated to improve the models. Furthermore, eventually, these predictive models may take on a prescriptive role and affect or optimize the course of action of a squadron. However, since resources can only be moved from one squadron to another, an enterprise-level algorithm will need to be created to determine the optimal allocation of resources. Lastly, a user-interface should be created so that a squadron's expected performance can be measured and tracked by maintainers, aviators, commanding officers, and leadership.

This page is intentionally blank.

REFERENCES

1. Larter, D. 2018. "The US Navy's fight to fix its worn-out Super Hornet fleet is making way," .
2. Insinna, V. 2018. "Boeing Super Hornet program gets second life through future sales and upgrades," .
3. Larter, D. 2018. "Flush with cash, the Navy bores in on aviation readiness amid a crisis," .
4. Mehta, A. 2018. "Mattis orders fighter jet readiness to jump to 80 percent - in one year," .
5. Blasius, D. T. and Doll, D. D. 2018. "The rise of man hours per flight hour," Tech. rep., CNA.
6. 2018. "List of United States Navy ratings," URL `https://en.wikipedia.org/wiki/List_of_United_States_Navy_ratings`.
7. Malone, L. M. J. 2018. "Understanding, Generating, and Calculating Capacity in Marine Aviation Squadrons," Tech. rep., Headquarters, United States Marine Corps.
8. Breiman, L., Friedman, J., Olshen, R., and Stone, C. 1984. "Classification and Regression Trees, 1984: Belmont," *CA: Wadsworth International Group*.
9. Meinshausen, N. 2006. "Quantile regression forests," *Journal of Machine Learning Research*, vol. 7, no. Jun, pp. 983–999.
10. Saabas, A. 2015. "Prediction intervals for Random Forests," Blog.
11. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., *et al.* 2011. "Scikit-learn: Machine learning in Python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830.
12. James, F. 1994. "MINUIT function minimization and error analysis: reference manual version 94.1," Tech. rep.
13. Brun, R. and Rademakers, F. 1997. "ROOT—an object oriented data analysis framework," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 389, no. 1-2, pp. 81–86.
14. Snedecor, G. W. and Cochran, W. G. 1989. "Statistical methods, 8th Edn," *Ames: Iowa State Univ. Press Iowa*.
15. Sasaki, Y. *et al.* 2007. "The truth of the F-measure," *Teach Tutor mater*, vol. 1, no. 5, pp. 1–5.
16. Pearson, K. 1896. "Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia," *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, vol. 187, pp. 253–318.
17. Powers, D. M. 2011. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," .
18. Wager, S., Hastie, T., and Efron, B. 2014. "Confidence intervals for random forests: The jackknife and the infinitesimal jackknife," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1625–1651.
19. Mentch, L. and Hooker, G. 2016. "Quantifying uncertainty in random forests via confidence intervals and hypothesis tests," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 841–881.

20. Collaboration, A. *et al.* 2015. "Modelling Z $\tau\tau$ processes in ATLAS with τ-embedded Z $\mu\mu$ data," *Journal of instrumentation*, vol. 10, no. 09, p. P09018.

APPENDIX A
EMBEDDED SIMULATION

Conceptually, embedded simulation [20] involves simulating event data where actual events are rare and getting more data is unfeasible. Embedded simulation is also contingent on two assumptions: the system is understood sufficiently well, and the event is independent of other activity in the sample or that dependence is known. Specifically, the rare event to be studied is simulated and embedded inside of actual data samples. In doing so, it is possible to create data that is realistic enough to make analysis decisions with or determine relationships in the data.

For the MC-FH model, it would be advantageous to know how levers interact and influence each other. Since many levers, such as pilot COB, are not expected to influence any other levers or drivers, that feature may be safely simulated and embedded in different samples. Usually, simulating a feature requires knowing the underlying distribution of that feature; however, in the case of levers like pilot COB, a simple grid search or choosing all possible values in a range is acceptable. By embedding the simulated features in different actual data samples, it is possible to generate a realistic larger dataset and study the MC-FH model response to samples with different combinations of lever values.

The features (levers) to be embedded and then simulated are:

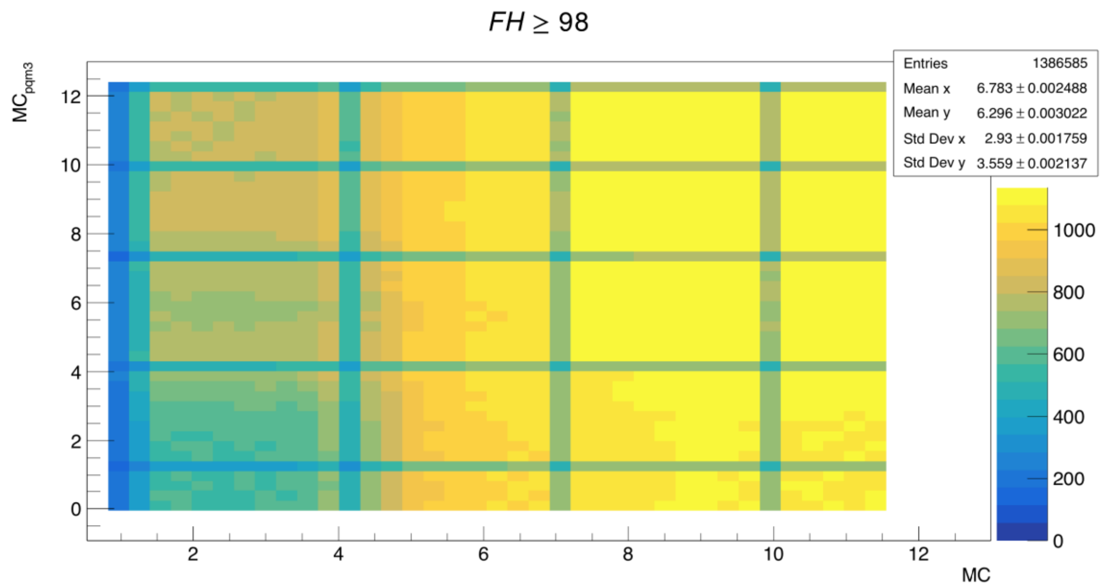
- $amsrr.mc\_mean\_fa - 18cef\_1$[4]
- $amsrr.mc\_mean\_fa - 18cef\_3\_pqm$
- $naeactuals.pilot\_cob\_1$.

MC and pilot COB values range between their $(min - 1)$ and $(max + 1)$ in step-sizes of 0.1. The hold-out set samples are then duplicated and the simulated levers are inserted. The total number of simulated samples is 1692600. Having a large number of simulated samples has the additional effect of reducing any biases that may be present in the actual data.

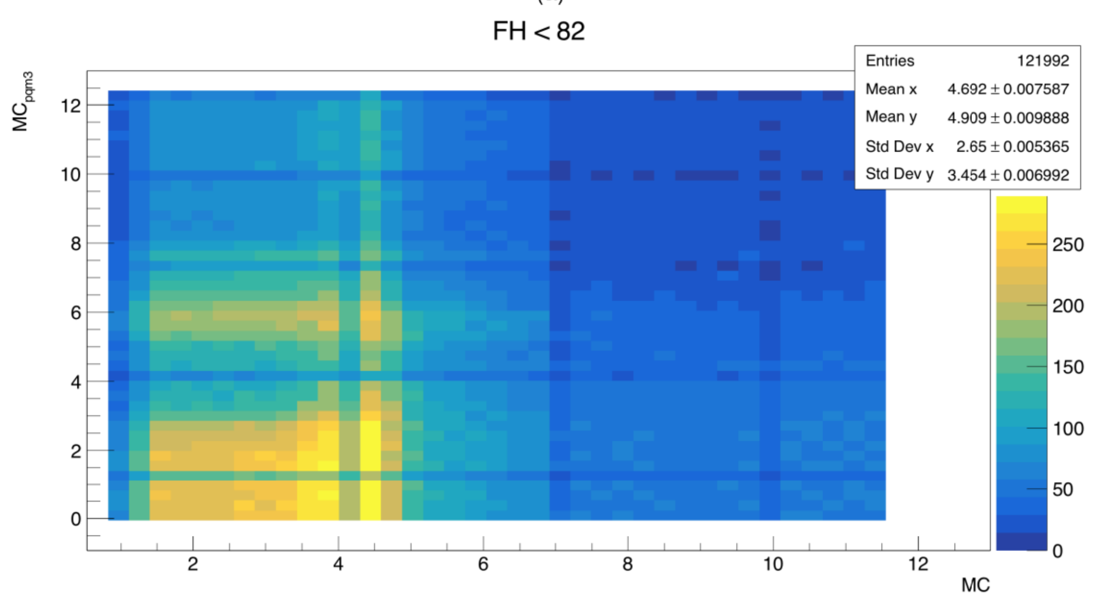
Figure A-1 shows 2-dimensional histograms of the simulated $amsrr.mc\_mean\_fa - 18cef\_3\_pqm$ versus the simulated $amsrr.mc\_mean\_fa - 18cef\_1$ as categorized by the month 1 MC-FH model. Figure A-1(a) shows the distribution of values when the sample is predicted by the month 1 MC-FH model to meet $\geq 98\%$ of their flight hour entitlement, and Figure A-1(b) shows the same distribution when the sample is predicted to meet $<82\%$ of their flight hour entitlement. It is clear that there is a distinct behavior in each plot where Figure A-1(b) has a higher concentration of samples in the lower left corner corresponding to a low MC_{pqm3} and a low MC, and Figure A-1(a) is more evenly distributed with a lower concentration of samples in the lower left corner. However, there is not a clean delineation between the plots and there is a non-zero background covering the space in Figure A-1(b) signifying that these two levers alone do not fully represent the MC-FH model. Figure A-2 shows only the normalized MC distributions from the embedded simulation and their classification according to the MC-FH model. Again, it is clear that MC is a good predictor, but the MC-FH model is much more complicated.

Additional parameters were studied in the same manner but were less human-interpretable.

[4]This is the feature predicted by the MC models.



(a)



(b)

Figure 20. (a) shows the relationship between MC_{pqm3} vs. MC determined via embedded simulation when the MC-FH model predicted ≥98% as a 2D histogram. (b) shows the relationship between MC_{pqm3} vs. MC determined via embedded simulation when the MC-FH model predicted <82% as a 2D histogram. Striations are due to binning choices and simulation step-size.

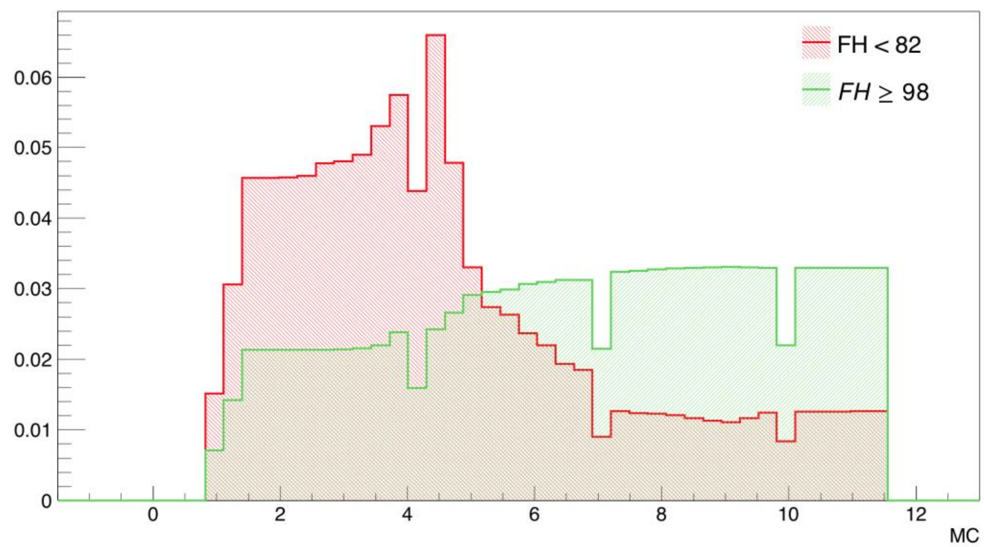


Figure 21. Normalized histogram of MC from embedded simulation classified according to MC-FH prediction.

INITIAL DISTRIBUTION

84310	Library	(1)
85300	Archive/Stock	(1)
54320	Dr. Benjamin Michlin	(1)
54340	Rick Cruz	(1)
71740	Josh Duclos	(1)
54320	Dean Lee	(1)
55180	Vincent Siu	(1)
56000	Charles Yetman	(1)

Defense Technical Information Center
Fort Belvoir, VA 22060–6218 (1)

REPORT DOCUMENTATION PAGE								Form Approved OMB No. 0704-01-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden to Department of Defense, Washington Headquarters Services Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*
December 2019

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE

Predicting FA-18 Squadron Readiness and
Quarterly Flight Hour Execution
Using Machine Learning

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHORS

Dr. Benjamin Michlin
Dr. Ruey Chang
Rick Cruz
Josh Duclos
Dean Lee
Vincent Siu
Charles Yetman
NIWC Pacific

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NIWC Pacific
53560 Hull Street
San Diego, CA 92152–5001

8. PERFORMING ORGANIZATION REPORT NUMBER

TR-3183

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Commander, Naval Air Forces (CNAF)
PO Box 357051
San Diego, CA 92135-7051

10. SPONSOR/MONITOR'S ACRONYM(S)

CNAF

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release.

13. SUPPLEMENTARY NOTES

This is a work of the United States Government and therefore is not copyrighted. This work may be copied and disseminated without restriction.

14. ABSTRACT

Given manning-training-equipment datasets from Naval FA-18 squadrons, a machine learning model for determining the monthly mean number of mission capable jets per squadron is created. This model is then extended and used as an input to create an ensemble of models determining the flight hour execution of a squadron over a three-month period. The ensemble of models is then used to predict squadron performance and readiness, and can correctly classify a squadron's future performance with 75% accuracy 90-days in advance.

15. SUBJECT TERMS

Data conditioning and transformation; data sources and aggregation; MC model creation and parameters; MC-FH Model; probabilistic interpretation; statistical interpretation; MC regression model; MC prediction residual; MC prediction interval; MC-FH confusion matrices; class probability vs MC ratio; embedded simulation for MC

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Benjamin Michlin
					19B. TELEPHONE NUMBER *(Include area code)*
U	U	U	U	50	619-553-2925

Standard Form 298 (Rev. 10/17)
Prescribed by ANSI Std. Z39.18

This page is intentionally blank.

This page is intentionally blank.

DISTRIBUTION STATEMENT A: Approved for public release.

Naval Information Warfare Center Pacific (NIWC Pacific)
San Diego, CA 92152-5001